```
Query: 181 EAAALYYLAAQHNVNALAMMTISDNLNNPEEDTSAEERQTTFTDMMKVGLETLISE 236
           EAAALYYLAAQH V+ALA+MTISD+L NP+EDT+AEERQ TFTDMMKVGLETLI++
Sbjct: 181 EAAALYYLAAQHQVDALAIMTISDSLVNPDEDTTAEERQNTFTDMMKVGLETLIAD 236
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2751> which encodes the amino acid sequence <SEQ ID 2752>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 210/235 (89%), Positives = 226/235 (95%)

Query: 1   MSIHIEAKQGEIADKILLPGDPLRAKFIAENFLEDAVCFNTVRNMFGYTGTYKGHRVSVM 60
            MSIHI AK+G+IADKILLPGDPLRAKFIAENFLEDAVCFN VRNMFGYTGTYKGHRVSVM
Sbjct: 1   MSIHISAKKGDIADKILLPGDPLRAKFIAENFLEDAVCFNEVRNMFGYTGTYKGHRVSVM 60

Query: 61  GTGMGMPSISIYARELIVDYGVKTLIRVGTAGAINPDIHVRELVLAQAAATNSNIIRNDW 120
            GTGMGMPSISIYARELIVDYGVKTLIRVGTAGAI+P++HVRELVLAQAAATNSNIIRND+
Sbjct: 61  GTGMGMPSISIYARELIVDYGVKTLIRVGTAGAIDPEVHVRELVLAQAAATNSNIIRNDF 120

Query: 121 PEFDFPQIADFKLLDKAYHIAKEMDITTHVGSVLSSDVFYSNQPDRNMALGKLGVHAIEM 180
            PEFDFPQIADF LLDKAYHIA+EM +TTHVG+VLSSDVFY+N P+RNMALGKLGV AIEM
Sbjct: 121 PEFDFPQIADFGLLDKAYHIAREMGVTTHVGNVLSSDVFYTNMPERNMALGKLGVKAIEM 180

Query: 181 EAAALYYLAAQHNVNALAMMTISDNLNNPEEDTSAEERQTTFTDMMKVGLETLIS 235
            EAAALYYLAAQH+V AL +MTISDNLN+P EDT+AEERQTTFTDMMKVGLETLI+
Sbjct: 181 EAAALYYLAAQHHVKALGIMTISDNLNDPTEDTTAEERQTTFTDMMKVGLETLIA 235
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 903

A DNA sequence (GBSx0958) was identified in *S.agalactiae* <SEQ ID 2753> which encodes the amino acid sequence <SEQ ID 2754>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1710(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9881> which encodes amino acid sequence <SEQ ID 9882> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2755> which encodes the amino acid sequence <SEQ ID 2756>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1386(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 126/253 (49%), Positives = 175/253 (68%), Gaps = 2/253 (0%)

Query: 3   IEMTDFSTALKVLVDQYSYHNAFLLLQKHGPLNSDLLFLLEMMKERRELNIDFLFAHQEQ 62
           + MT+  T L +L+D Y+Y++AF + +    +    L+LLEM+KERRELN+ FL  H  +
Sbjct: 1   LPMTNNQT-LDILLDVYAYNHAFRIAKALPNIPKTALYLLEMLKERRELNLAFLAEHAAE 59

Query: 63  VVILQEKYNIKL-LHNPYDLELLANYIMDLEAKVKNGLIIDFVRSVSPILYRLFMILLAQ 121
              ++++Y+  L L+   + E +ANYI+DLE KVKNG IIDFVRSVSPILYRLF+ L+
Sbjct: 60  NRTIEDQYHCSLWLNQSLEDEQIANYILDLEVKVKNGAIIDFVRSVSPILYRLFLRLITS 119

Query: 122 EVPHLHDYIHNARDDHYDTWKFKELKESNHPVLLAFSERWHDSRLTSKSLAECLQLTDLD 181
           E+P+   YI + ++D YDTW F+ + ES+H V  A+  +     +T+KSLA+ L LT L
Sbjct: 120 EIPNFKAYIFDTKNDQYDTWHFQAMLESDHEVFKAYLSQKQSRNVTTKSLADMLTLTSLP 179

Query: 182 EEVKSTIIQLRQFEKSVRNPLAHLIKPFDEQELYRTTQFSSQAFLDQIIFLAKVIGVEYD 241
           +E+K  +  LR FEK+VRNPLAHLIKPFDE+EL+RTT FSSQAFL+ II LA   GV Y
Sbjct: 180 QEIKDLVFLLRHFEKAVRNPLAHLIKPFDEEELHRTTHFSSQAFLENIITLATFSGVIYR 239

Query: 242 TVNFHYDTVNKLI 254
              F++D +N +I
Sbjct: 240 REPFYFDDMNAII 252
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 904

A DNA sequence (GBSx0959) was identified in *S.agalactiae* <SEQ ID 2757> which encodes the amino acid sequence <SEQ ID 2758>. This protein is predicted to be CpsY protein. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane  260 - 276 ( 260 - 276)

----- Final Results -----
              bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9879> which encodes amino acid sequence <SEQ ID 9880> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2759> which encodes the amino acid sequence <SEQ ID 2760>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1958(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 247/301 (82%), Positives = 274/301 (90%)

Query: 1   MRIQQLQYVIKIVETGSMNEAAKQLYITQPSLSNAVRNLETEMGIQIFIRNPKGITLTKD 60
           MRIQQL Y+IKIVE GSMNEAAKQL+ITQPSLSNAV++LE EMGI IF RNPKGITLTKD
Sbjct: 1   MRIQQLHYIIKIVECGSMNEAAKQLFITQPSLSNAVKDLEMEMGITIFNRNPKGITLTKD 60

Query: 61  GMEFLSYARQILEQTALLEERYKGDNTSRELFSVSSQHYAFVVNAFVALFNGTDMTQYEL 120
           G+EFLSYARQI+EQT+LLE+RYK  NT RELFSVSSQHYAFVVNAFV+L   TDMT+YEL
Sbjct: 61  GVEFLSYARQIIEQTSLLEDRYKNHNTGRELFSVSSQHYAFVVNAFVSLLKRTDMTRYEL 120

Query: 121 FLRETRTWEIIDDVKNFRSEIGVLFLNSYNRDVLTKLFDDNSLIATTLFTTTPHIFVSKS 180
           FLRETRTWEIIDDVKNFRSEIGVLF+N YNRDVLTKLFDDN L A+ LF   PHIFVSKS
Sbjct: 121 FLRETRTWEIIDDVKNFRSEIGVLFINDYNRDVLTKLFDDNHLTASPLFKAQPHIFVSKS 180

Query: 181 NPLANRKKLNMKDLEDYPYLSYDQGLHNSFYFSEEMMSQIPHPKSIVVSDRATLFNLMIG 240
           NPLA +  L+M DL D+PYLSYDQG+HNSFYFSEEMMSQ+PH KSIVVSDRATLFNLMIG
Sbjct: 181 NPLATKSLLSMDDLRDFPYLSYDQGIHNSFYFSEEMMSQMPHNKSIVVSDRATLFNLMIG 240

Query: 241 LDGYTVATGILNSKLNGDEIVAIPLDVDDVIDIVYIRHDKANLSKMGQKFIDYLLEEVSFN 301
           LDGYTVA+GILNS LNGD+IVAIPLDV D IDIV+I+H+KANLSKMG++FI+YLLEEV+F+
Sbjct: 241 LDGYTVASGILNSNLNGDQIVAIPLDVPDEIDIVFIKHEKANLSKMGERFIEYLLEEVTFD 301
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 905

A DNA sequence (GBSx0960) was identified in *S.agalactiae* <SEQ ID 2761> which encodes the amino acid sequence <SEQ ID 2762>. This protein is predicted to be CpsX protein. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.91   Transmembrane    22 -  38 (  13 -  42)
   INTEGRAL    Likelihood =-14.65   Transmembrane    52 -  68 (  44 -  77)
   INTEGRAL    Likelihood = -6.74   Transmembrane    76 -  92 (  73 -  97)

----- Final Results -----
               bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC44935 GB:U56901 putative transcriptional regulator [Bacillus subtilis]
 Identities = 120/389 (30%), Positives = 196/389 (49%), Gaps = 17/389 (4%)

Query: 2   KIGKKIVLMFTAIVLTTVLALGVYLTSAYTFSTGELSKTFKDFSTSSNKSDAIK-QTRAF 60
           KI K+I+L+F A+ L  V+ LG Y       +  E        + S+ +++ +  + + F
Sbjct: 19  KILKRIMLLF-ALALLVVVGLGGYKLYKTINAADESYDALSRGNKSNLRNEVVDMKKKPF 77

Query: 61  SILLMGVDTGSSERASKWEGNSDSMILVTVNPKTKKTTMTSLERDTLTTLSGPKNNEMNG 120
           SIL MG++  +++     +G SDS+I+VT++PK K   M S+ RDT   L+G    +  G
Sbjct: 78  SILFMGIEDYATKGQ---KGRSDSLIVVTLDPKNKTMKMLSIPRDTRVQLAG----DTTG 130

Query: 121 VEAKLNAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGLIDLVNAVGGITVTNEFDFPI 180
            + K+NAAY+ GG    + TV++ L I ID YV ++  G  D++N VGGI V   FDF
Sbjct: 131 SKTKINAAYSKGGKDETVETVENFLQIPIDKYVTVDFDGFKDVINEVGGIDVDVPFDFDE 190

Query: 181 SIAENEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240
               +E + +     G   +NGE+AL YARMR  D  GD+GR  RQ++++  ++ ++ +
Sbjct: 191 KSDVDESK-RIYFKKGEMHLNGEEALAYARMRKQDKRGDFGRNDRQKQILNALIDRMSSA 249

Query: 241 DSISSYRKILSAVSSNMQTNIEISSRTIPSLLGYRDALRTIKTYQLKGEDATLSDGGSYQ 300
            +I+   KI    S N++TNI I+       +      + I T  + G D  L    +Y
```

```
Sbjct: 250 SNIAKIDKIAEKASENVETNIRITEGLALQQIYSGFTSKKIDTLSITGSDLYLGPNNTYY 309

Query: 301 IVTSNHLLEIQNRIRTELGLHKVNQLKTNATVYENLYGSTKSQTVNNNYDSSGQAPSYSD 360
                   LE   ++R  L  H ++      +T         T S + + +  S+G      +
Sbjct: 310 FEPDATNLE---KVRKTLQEH-LDYTPDTSTGTSGTEDGTDSSSSSGSTGSTGTTTDGTT 365

Query: 361 SHSSYANYSSGVDTGQSASTDQDSTASSH 389
            + SSY+N SS    T  + ST   +T SS+
Sbjct: 366 NGSSYSNDSS---TSSNNSTTNSTTDSSY 391
```

There is also homology to SEQ ID 2764.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 906

A DNA sequence (GBSx0961) was identified in *S.agalactiae* <SEQ ID 2765> which encodes the amino acid sequence <SEQ ID 2766>. This protein is predicted to be CpsIaB. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  121 - 137 ( 121 - 137)

----- Final Results -----
              bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9877> which encodes amino acid sequence <SEQ ID 9878> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 907

A DNA sequence (GBSx0962) was identified in *S.agalactiae* <SEQ ID 2767> which encodes the amino acid sequence <SEQ ID 2768>. This protein is predicted to be cpsb protein. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.02   Transmembrane  182 - 198 ( 179 - 204)
   INTEGRAL    Likelihood = -5.57   Transmembrane   30 -  46 (  24 -  48)

----- Final Results -----
              bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 10785> and protein <SEQ ID 10786> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
```

```
McG: Discrim Score:     -8.96
GvH: Signal Score (-7.5): 0.11
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 2 value:  -9.02 threshold:  0.0
   INTEGRAL    Likelihood = -9.02   Transmembrane  182 - 198 ( 179 - 204)
   INTEGRAL    Likelihood = -5.57   Transmembrane   30 -  46 (  24 -  48)
   PERIPHERAL  Likelihood =  6.21     113
 modified ALOM score:   2.30

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 908

A DNA sequence (GBSx0963) was identified in *S.agalactiae* <SEQ ID 2769> which encodes the amino acid sequence <SEQ ID 2770>. This protein is predicted to be CpsIaD. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane  149 - 165 ( 149 - 166)

----- Final Results -----
              bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 909

A DNA sequence (GBSx0964) was identified in *S.agalactiae* <SEQ ID 2771> which encodes the amino acid sequence <SEQ ID 2772>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.26   Transmembrane  276 - 292 ( 270 - 297)
   INTEGRAL    Likelihood = -4.62   Transmembrane   10 -  26 (   9 -  28)
   INTEGRAL    Likelihood = -4.14   Transmembrane   41 -  57 (  39 -  58)
   INTEGRAL    Likelihood = -3.24   Transmembrane  100 - 116 ( 100 - 116)
   INTEGRAL    Likelihood = -3.08   Transmembrane  445 - 461 ( 443 - 461)

----- Final Results -----
              bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8687> and protein <SEQ ID 8688> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      5.69
GvH: Signal Score (-7.5): -5.63
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value: -12.26 threshold:  0.0
   INTEGRAL    Likelihood =-12.26   Transmembrane  276 - 292 ( 270 - 297)
   INTEGRAL    Likelihood = -4.62   Transmembrane   10 -  26 (   9 -  28)
   INTEGRAL    Likelihood = -4.14   Transmembrane   41 -  57 (  39 -  58)
   INTEGRAL    Likelihood = -3.24   Transmembrane  100 - 116 ( 100 - 116)
   INTEGRAL    Likelihood = -3.08   Transmembrane  445 - 461 ( 443 - 461)
   PERIPHERAL  Likelihood =  2.23     221
 modified ALOM score:   2.95

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 910

A DNA sequence (GBSx0965) was identified in *S.agalactiae* <SEQ ID 2773> which encodes the amino acid sequence <SEQ ID 2774>. This protein is predicted to be CpsF. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane   79 -  95 (  78 -  95)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 911

A DNA sequence (GBSx0966) was identified in *S.agalactiae* <SEQ ID 2775> which encodes the amino acid sequence <SEQ ID 2776>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4634(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 912

A DNA sequence (GBSx0967) was identified in *S.agalactiae* <SEQ ID 2777> which encodes the amino acid sequence <SEQ ID 2778>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.47   Transmembrane   59 -  75 (  54 -  82)
   INTEGRAL    Likelihood =-10.88   Transmembrane  309 - 325 ( 307 - 332)
   INTEGRAL    Likelihood = -8.07   Transmembrane   33 -  49 (  28 -  53)
   INTEGRAL    Likelihood = -6.48   Transmembrane  195 - 211 ( 187 - 212)
   INTEGRAL    Likelihood = -6.16   Transmembrane  285 - 301 ( 283 - 306)
   INTEGRAL    Likelihood = -4.09   Transmembrane  222 - 238 ( 221 - 240)
   INTEGRAL    Likelihood = -3.50   Transmembrane   78 -  94 (  77 -  96)
   INTEGRAL    Likelihood = -2.71   Transmembrane  101 - 117 (  99 - 117)
   INTEGRAL    Likelihood = -2.44   Transmembrane    8 -  24 (   7 -  25)
   INTEGRAL    Likelihood = -1.59   Transmembrane  147 - 163 ( 147 - 164)
   INTEGRAL    Likelihood = -0.48   Transmembrane  168 - 184 ( 168 - 184)

----- Final Results -----
               bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB43614 GB:AJ239004 polysaccharide polymerase [Streptococcus pneumoniae]
 Identities = 74/309 (23%), Positives = 137/309 (43%), Gaps = 36/309 (11%)

Query: 53  FERRKLV---IIFLLFIATILNLFFVHKVTFILTLIFFLALKDI--SLKKAFSIIIGSRI 107
           FE+RK     II ++ I T+L    +    ++    +F+ +  I   L++   II
Sbjct: 61  FEKRKYTLQFIISIILITTLLLYTSIQMQNYVYFTSWFMLIGTIHYDLRRVIKIIFIVS- 119

Query: 108 LGVLLNQIFVKLDLIEIKY-----VNFYRDGQFILRSDLGFGHPNFIHNFFALTIFLYIV 162
           L ++   IF+ L  + I Y     +N  R+ + +     GF HPN      +    ++I
Sbjct: 120 LSIMFISIFISLLMYIIDYKREILINIRRN-ETVRAFTFGFIHPNKFTIVLSNLCLMFIW 178

Query: 163 LNYKRLKPVVMVLFLTLNYLLYQYTFSRTGYYIVILFIVLIYVTKNSLIKRVFMKLAPYV 222
           L   RLK   +   L +    Y +T +RT   + I+   L+Y+     ++ + ++   Y
Sbjct: 179 LIKDRLKYYHVTFCLFIQLFFYFFTQTRTALLVSIVIFALLYI--YMFVENLELRWIGYS 236

Query: 223 QFFLLVFTFLSSTIFFNSN--FVQKLDVLLTGRLHY-AHLQLVDGLTPFGNSFKE----- 274
            F +  F  + +  F+ SN  F   +D +LTGR+   A+ +   G T +G    +
Sbjct: 237 FFCISTFLGVLAFQFYPSNNKFSIFIDNILTGRIKLAAYARTFFGYTFWGQYVDKEIVWD 296

Query: 275 -----TSVLFDNSYSMLLSMYGVVLTMFCMIIY-----YIYSKKIIIIELQLLLFIMSII 324
                TS  FD+ YS L+S  G++  +   +++     Y+ +K +I+    LL + M  +
Sbjct: 297 PIWGLTSFTFDSFYSFLMSNAGIIWLLILSVLFVKLQKYLDNKSLIL----LLAWSMYAV 352

Query: 325 LFTESFYPS 333
             T+  +PS
Sbjct: 353 TETDLIFPS 361
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 913

A DNA sequence (GBSx0968) was identified in *S.agalactiae* <SEQ ID 2779> which encodes the amino acid sequence <SEQ ID 2780>. This protein is predicted to be cap8J. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3424(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB43613 GB:AJ239004 cap8J [Streptococcus pneumoniae]
 Identities = 94/237 (39%), Positives = 135/237 (56%), Gaps = 10/237 (4%)

Query: 1    MIPKVIHYCWFGGNPLPDNLKKYIKTWREQCPDYEIIEWNEHNYDVSKNVFMREAYTKKN 60
            MIPK IHY WFGG+  PD + K I +W++  PDYEI+EWNE N+D+S + F + AY  +
Sbjct: 1    MIPKKIHYIWFGGSEKPDVVLKCINSWKKYMPDYEIVEWNEDNFDLSDSQFAKSAYESRK 60

Query: 61   FAYVSDYARLDIIYTYGGFYLDTDVELLKSL-DPLRIHECFLAREISCDVNTGLIIGAVK 119
            +A+ SDYAR  I+  YGG Y DTDVELLK++ D +  H  F   E   +VN GL+    +
Sbjct: 61   WAFASDYARFKILSKYGGIYFDTDVELLKTISDDILAHSSFTGFEYIGEVNPGLVYACMP 120

Query: 120 GHHFLKSNMSIYDKS--DLTSLNKTCVEVTTNLLINRGLKNKNIIQKIDDITIYPRNYFN 177
                 K  +  Y+++  D+  L  T   + T+ L+    +  N  Q ID + IYP +YF
Sbjct: 121 DDKIAKYMVQYYEQASFDINHL-VTVNTIITDYLLKNNFQKNNQFQIIDGLAIYPDDYFC 179

Query: 178 PKNLLTGKVDCLTSVTYSIHHYEGSWKSSSFISDSLKIRVRLIIDFLFGYGTYRMLL 234
              +    +V  LT  T SIHHY  +WK+      +LK +V++I+  + G   YR LL
Sbjct: 180 GYDQEVKEVR-LTERTISIHHYSATWKTR-----TLKRKVQMIVKTIIGAENYRKLL 230
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 914

A DNA sequence (GBSx0969) was identified in *S.agalactiae* <SEQ ID 2781> which encodes the amino acid sequence <SEQ ID 2782>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3897(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA87700 GB:Z47767 WbcL [Yersinia enterocolitica]
 Identities = 60/207 (28%), Positives = 101/207 (47%), Gaps = 22/207 (10%)

Query: 4    IFTPTFNRGYRLSYLYDSLCNQTNKNFIWLIVDDGSEDSTKEIVSNYIKENKVSIVYLYK 63
            +FTPTFNR + L   Y S+  Q   +  WLIVDDGS D+T E+V ++  ENK++I Y+Y+
Sbjct: 6    VFTPTFNRAHVLKRCYLSILEQDRDDIEWLIVDDGSTDNTAEVVDSFKIENKLNIKYIYQ 65

Query: 64   RNGGKHSAYNLAMRYMQPSDYHVCVDSDDWLLEDAV------EIIFKDLESLTLSNRYVG 117
             N GK +A+N A+      +Y + +DSDD  +  ++        +F D E + +
```

```
Sbjct: 66  DNSGKQAAWNKAVENAS-GEYFIGLDSDDAFIAGSINKLLSMNAVFDDKEIIGIR----A 120

Query: 118 LVYPRYSLNQGNNWLNPKILEVNIPDLKYKYHLKIETCIVINNAYLVDFEFPCFEGENFL 177
           +      +L   N +L+ +  + +  D ++   ++ E         L  + +P   G NF+
Sbjct: 121 ISVSSETLKPNNYYLSNEDKKSSWFD-EFSSGIRGERIDFFKTELLRKYLYPVASGINFI 179

Query: 178 SEEIMYIYLSKKGYFCPQNRKIYCFDY 204
            E   Y  ++K+          YCF Y
Sbjct: 180 PEIWFYSTVAKE----------YCFYY 196
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 915

A DNA sequence (GBSx0970) was identified in *S.agalactiae* <SEQ ID 2783> which encodes the amino acid sequence <SEQ ID 2784>. This protein is predicted to be eps7. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18    Transmembrane  190 - 206 ( 189 - 206)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59293 GB:AJ131984 putative galactosyl transferase
           [Streptococcus pneumoniae]
 Identities = 101/312 (32%), Positives = 172/312 (54%), Gaps = 4/312 (1%)

Query: 3   LISIIVPVYNGEIYIGRCLDSILEQTYQNLEIIIIDDGSSDRTGDICEKYFLEDRRIKYF 62
           +IS+IVPVYN   Y+   LDS+LEQTY++ E+I+++DGS+D +G+IC++Y      I  F
Sbjct: 1   MISVIVPVYNVADYLRFALDSLLEQTYKDFEVILVNDGSTDNSGEICDEYGKLYDNIHVF 60

Query: 63  YQENRGQSVARNNGVLRCTGDWIAFLDSDDVYLPYSIEVMYNIQKATNADIVLT--SIGN 120
           +++N G S ARN G+ +  G++I FLDSDD + PY++E++  IQK  + DIV T   I
Sbjct: 61  HKKNGGLSDARNFGLEKSRGEFITFLDSDDYFEPYALELLITIQKKYDVDIVSTKGGITY 120

Query: 121 FNNTYNTSINSQYLKEIKLYTLEVALEEMYYGKTYGVSPLAKLYPRSNLLSNPYPEGKIH 180
            ++ Y+  + ++    +K+ T +  L  +YY     VS   KLY R +L    +P+GKI+
Sbjct: 121 SHDIYSKKLMAEDYLTVKILTNKEFLAAVYYNDEMTVSAWGKLYKR-DLFKTIFPKGKIY 179

Query: 181 EDMDTTFKLISCASKIAVCDIVTAVVYFSDNSTTRTKFNERMLYFFEAIQNNIVFINLNF 240
           ED+    + +     +A  D+     Y    S   + F++R   FF+AI +N   I   +
Sbjct: 180 EDLYVVAERLLNIKTVAHTDLPIYHYYQRQGSIVNSTFSDRQYDFFDAIDHNEAIIKKFY 239

Query: 241 PHNTSLISAVIYNEVFGGIDICGKMIDFKLYDTVDYYRKKYRKYFKTILFNNRISVKEKV 300
             +  L++A+    V G   I      +   + +    +  + Y+  ++ N +I +K KV
Sbjct: 240 CGDKELLAALNAKRVIGSF-ILSNSAFYNSKNDITKIIRIIKPYYWEVIKNKKIPMKRKV 298

Query: 301 KYILFISSIRYF 312
           + +LF+ S  Y+
Sbjct: 299 QCVLFLLSPNYY 310
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 916

A DNA sequence (GBSx0971) was identified in *S.agalactiae* <SEQ ID 2785> which encodes the amino acid sequence <SEQ ID 2786>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2787> which encodes the amino acid sequence <SEQ ID 2788>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2065(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 37/111 (33%), Positives = 61/111 (54%), Gaps = 3/111 (2%)

Query: 1    MDKVSIIIPVYNVQSFLNECIESVLAQ-TYSNLEIILVNDGSTDNSGDIC-DYYSEIDGR 58
            M KVSII   YN   ++++ ++S L+Q T   +EII+++D STD+S +I   Y +   G+
Sbjct: 1    MYKVSIICTNYNKAPWISDALDSFLSQVTDFEVEIIVIDDASTDDSREILKSYQKKSSGK 60

Query: 59   I-FVFHKNNGGLSDARNYGISRATGDYIYLLDSDDYLYKEDAIERMVEFSE 108
            I  +F++ N G++         A G YI   D DDY      +++ V+  E
Sbjct: 61   IKLLFNETNIGITKTWIKACLYAKGKYIARCDGDDYWTDSFKLQKQVDVLE 111
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 917

A DNA sequence (GBSx0972) was identified in *S.agalactiae* <SEQ ID 2789> which encodes the amino acid sequence <SEQ ID 2790>. This protein is predicted to be CpsK. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 918

A DNA sequence (GBSx0973) was identified in *S.agalactiae* <SEQ ID 2791> which encodes the amino acid sequence <SEQ ID 2792>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1956(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 919

A DNA sequence (GBSx0974) was identified in *S.agalactiae* <SEQ ID 2793> which encodes the amino acid sequence <SEQ ID 2794>. This protein is predicted to be capsular polysaccharide. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.81   Transmembrane   89 - 105 (  80 - 112)
   INTEGRAL    Likelihood = -7.01   Transmembrane  439 - 455 ( 428 - 460)
   INTEGRAL    Likelihood = -6.74   Transmembrane  322 - 338 ( 317 - 342)
   INTEGRAL    Likelihood = -4.88   Transmembrane  175 - 191 ( 174 - 195)
   INTEGRAL    Likelihood = -3.45   Transmembrane  146 - 162 ( 145 - 166)
   INTEGRAL    Likelihood = -3.08   Transmembrane  381 - 397 ( 375 - 398)
   INTEGRAL    Likelihood = -2.50   Transmembrane  413 - 429 ( 412 - 430)
   INTEGRAL    Likelihood = -1.91   Transmembrane  206 - 222 ( 205 - 222)
   INTEGRAL    Likelihood = -1.59   Transmembrane  354 - 370 ( 354 - 372)
   INTEGRAL    Likelihood = -1.54   Transmembrane   43 -  59 (  43 -  61)
   INTEGRAL    Likelihood = -0.22   Transmembrane  252 - 268 ( 252 - 268)

----- Final Results -----
               bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 920

A DNA sequence (GBSx0975) was identified in *S.agalactiae* <SEQ ID 2795> which encodes the amino acid sequence <SEQ ID 2796>. This protein is predicted to be NeuB. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2992(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 921

A DNA sequence (GBSx0976) was identified in *S.agalactiae* <SEQ ID 2797> which encodes the amino acid sequence <SEQ ID 2798>. This protein is predicted to be NeuC. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3150(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 922

A DNA sequence (GBSx0977) was identified in *S.agalactiae* <SEQ ID 2799> which encodes the amino acid sequence <SEQ ID 2800>. This protein is predicted to be neuD. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

There is homology to SEQ ID 542.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 923

A DNA sequence (GBSx0979) was identified in *S.agalactiae* <SEQ ID 2801> which encodes the amino acid sequence <SEQ ID 2802>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 924

A DNA sequence (GBSx0980) was identified in *S.agalactiae* <SEQ ID 2803> which encodes the amino acid sequence <SEQ ID 2804>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1621(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9875> which encodes amino acid sequence <SEQ ID 9876> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2805> which encodes the amino acid sequence <SEQ ID 2806>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1066(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 83/139 (59%), Positives = 111/139 (79%)

Query: 6   TETHDHQALIQKLLVSIHYLTLFRDEIILVEKTPSLLGKHFSIAIVQNELGEILSKIEAL 65
           TE + HQ LIQKLLVSIHYLTLFRDE+ LVE+TPS+LG  F   +VQ+ELG+I++ I+ L
Sbjct: 4   TEQNSHQILIQKLLVSIHYLTLFRDELKLVERTPSILGGEFPAHLVQSELGDIVAAIDTL 63

Query: 66  SKQKKLIRSIYWYDESSFKVMNKALAIVEEWIKGLDNLLEFCQSQTVFQAILGDERAHVF 125
             Q++LI S +WY+ES+FK+MNK L IV+ WIKG+D+L++ CQS+ VFQ I+GD+R  VF
Sbjct: 64  DMQQRLIESTFWYEESAFKLMNKTLDIVDNWIKGVDHLIDLCQSKEVFQIIIGDKRIRVF 123

Query: 126 GILIDVYTSLNIINTSLKE 144
           G+L DV++SL +   SLKE
Sbjct: 124 GVLSDVFSSLKVSALSLKE 142
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 925

A DNA sequence (GBSx0981) was identified in *S.agalactiae* <SEQ ID 2807> which encodes the amino acid sequence <SEQ ID 2808>. This protein is predicted to be uracil-DNA glycosylase (ung). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3427(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2809> which encodes the amino acid sequence <SEQ ID 2810>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4200(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 160/216 (74%), Positives = 185/216 (85%)

Query: 1   MKHSSWHDLIKRELPNHYYNKINTFMDAVYESGIVYPPRDKVFNAIQITPLENVKVVIIG 60
           M HS WH+ IK  LP HYY +IN F+D  Y SG+VYPPR+ VF A+Q+TPLE  KV+I+G
Sbjct: 1   MAHSIWHEKIKSFLPEHYYGRINHFLDEAYASGLVYPPRENVFKALQVTPLEETKVLILG 60

Query: 61  QDPYHGPQQAQGLSFSVPDNLPAPPSLQNILKELAEDIGSRSHHDLTSWAQQGVLLLNAC 120
           QDPYHGP+QAQGLSFSVP+ + APPSL NILKELA+DIG R HHDL++WA QGVLLLNAC
Sbjct: 61  QDPYHGPKQAQGLSFSVPEEISAPPSLINILKELADDIGPRDHHDLSTWASQGVLLLNAC 120

Query: 121 LTVPEHQANGHAGLIWEPFTDAVIKVVNQKETPVVFILWGGYARKKKSLIDNPIHHIIES 180
           LTVP  QANGHAGLIWEPFTDAVIKV+N+K++PVVFILWG YARKKK+ I NP HHIIES
Sbjct: 121 LTVPAGQANGHAGLIWEPFTDAVIKVLNEKDSPVVFILWGAYARKKKAFITNPKHHIIES 180

Query: 181 PHPSPLSAYRGFFGSRPFSRTNHFLEEEGINEIDWL 216
           PHPSPLS+YRGFFGS+PFSRTN  LE+EG+   +DWL
Sbjct: 181 PHPSPLSSYRGFFGSKPFSRTNAILEKEGMTGVDWL 216
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 926

A DNA sequence (GBSx0982) was identified in *S.agalactiae* <SEQ ID 2811> which encodes the amino acid sequence <SEQ ID 2812>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.15   Transmembrane  147 - 163 ( 109 - 166)
   INTEGRAL    Likelihood = -8.92   Transmembrane  124 - 140 ( 109 - 146)
   INTEGRAL    Likelihood = -6.16   Transmembrane  167 - 183 ( 166 - 186)
   INTEGRAL    Likelihood = -4.67   Transmembrane    3 -  19 (   1 -  23)
   INTEGRAL    Likelihood = -3.98   Transmembrane   72 -  88 (  64 -  92)
   INTEGRAL    Likelihood = -1.06   Transmembrane  106 - 122 ( 105 - 122)
   INTEGRAL    Likelihood = -0.90   Transmembrane   54 -  70 (  54 -  70)

----- Final Results -----
               bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9873> which encodes amino acid sequence <SEQ ID 9874> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA91549 GB:Z67739 unidentified [Streptococcus pneumoniae]
 Identities = 134/212 (63%), Positives = 168/212 (79%)

Query: 1   MNIIIMIIIAYLLGSIQTGLWIGKYFYQVNLRQHGSGNTGTTNTFRILGVKAGIVTLTID 60
```

```
           M  I+++I+AYLLGSI +GLWIG+ F+Q+NLR+HGSGNTGTTNTFRILG KAG+ T  ID
Sbjct: 1   MITIVLLILAYLLGSIPSGLWIGQVFFQINLREHGSGNTGTTNTFRILGKKAGMATFVID 60

Query: 61  ILKGTLATLIPIILGITTVSPFFIGFFAIIGHTFPIFAQFKGGKAVATSAGVLLGFAPSF 120
              KGTLATL+PII  +  VSP   G  A+IGHTFPIFA FKGGKAVATSAGV+ GFAP F
Sbjct: 61  FFKGTLATLLPIIFHLQGVSPLIFGLLAVIGHTFPIFAGFKGGKAVATSAGVIFGFAPIF 120

Query: 121 FLYLLVIFLLTLYLFSMISLSSITVAVVGILSVLIFPLVGFILTDYDWIFTTVVILMALT 180
             LYL +IF   LYL SMISLSS+T ++  ++ VL+FPL GFIL++YD++F  +++ +A
Sbjct: 121 CLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDFLFIAIILALASL 180

Query: 181 IIIRHQDNIKRIRKRQENLVPFGLNLSKQKNK 212
            IIIRH+DNI RI+ + ENLVP+GLNL+ Q  K
Sbjct: 181 IIIRHKDNIARIKNKTENLVPWGLNLTHQDPK 212
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2813> which encodes the amino acid sequence <SEQ ID 2814>. Analysis of this protein sequence reveals the following:

```
      Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.83   Transmembrane  194 - 210 ( 191 - 216)
   INTEGRAL    Likelihood = -9.77   Transmembrane  146 - 162 ( 132 - 191)
   INTEGRAL    Likelihood = -7.70   Transmembrane  165 - 181 ( 163 - 191)
   INTEGRAL    Likelihood = -5.89   Transmembrane   23 -  39 (  19 -  47)
   INTEGRAL    Likelihood = -4.83   Transmembrane   95 - 111 (  91 - 118)

----- Final Results -----
               bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA91549 GB:Z67739 unidentified [Streptococcus pneumoniae]
 Identities = 138/213 (64%), Positives = 166/213 (77%)

Query: 28  MKLLLFITIAYLLGSIPTGLWIGQYFYHINLREHGSGNTGTTNTFRILGVKAGTATLAID 87
            M  ++ + +AYLLGSIP+GLWIGQ F+ INLREHGSGNTGTTNTFRILG KAG AT  ID
Sbjct: 1   MITIVLLILAYLLGSIPSGLWIGQVFFQINLREHGSGNTGTTNTFRILGKKAGMATFVID 60

Query: 88  MFKGTLSILLPIIFGMTSISSIAIGFFAVLGHTFPIFANFKGGKAVATSAGVLLGFAPLY 147
             FKGTL+ LLPIIF +  +S +  G  AV+GHTFPIFA FKGGKAVATSAGV+ GFAP++
Sbjct: 61  FFKGTLATLLPIIFHLQGVSPLIFGLLAVIGHTFPIFAGFKGGKAVATSAGVIFGFAPIF 120

Query: 148 LFFLASIFVLVLYLFSMISLASVVSAIVGVLSVLTFPAIHFLLPNYDYFLTFIVILLAFI 207
              +LA IF   LYL SMISL+SV ++I  V+ VL FP   F+L NYD+    I++ LA +
Sbjct: 121 CLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDFLFIAIILALASL 180

Query: 208 IIIRHKDNISRIKHHTENLIPWGLNLSKQVPKK 240
            IIIRHKDNI+RIK+ TENL+PWGLNL+ Q PKK
Sbjct: 181 IIIRHKDNIARIKNKTENLVPWGLNLTHQDPKK 213
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 143/212 (67%), Positives = 174/212 (81%)

Query: 1   MNIIIMIIIAYLLGSIQTGLWIGKYFYQVNLRQHGSGNTGTTNTFRILGVKAGIVTLTID 60
            M +++ I IAYLLGSI TGLWIG+YFY +NLR+HGSGNTGTTNTFRILGVKAG  TL ID
Sbjct: 28  MKLLLFITIAYLLGSIPTGLWIGQYFYHINLREHGSGNTGTTNTFRILGVKAGTATLAID 87

Query: 61  ILKGTLATLIPIILGITTVSPFFIGFFAIIGHTFPIFAQFKGGKAVATSAGVLLGFAPSF 120
            + KGTL+ L+PII G+T++S   IGFFA++GHTFPIFA FKGGKAVATSAGVLLGFAP +
Sbjct: 88  MFKGTLSILLPIIFGMTSISSIAIGFFAVLGHTFPIFANFKGGKAVATSAGVLLGFAPLY 147

Query: 121 FLYLLVIFLLTLYLFSMISLSSITVAVVGILSVLIFPLVGFILTDYDWIFTTVVILMALT 180
              +L  IF+L LYLFSMISL+S+  A+VG+LSVL FP + F+L +YD+  T +VIL+A
Sbjct: 148 LFFLASIFVLVLYLFSMISLASVVSAIVGVLSVLTFPAIHFLLPNYDYFLTFIVILLAFI 207
```

```
Query: 181 IIIRHQDNIKRIRKRQENLVPFGLNLSKQKNK 212
           IIIRH+DNI RI+   ENL+P+GLNLSKQ  K
Sbjct: 208 IIIRHKDNISRIKHHTENLIPWGLNLSKQVPK 239
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 927

A DNA sequence (GBSx0983) was identified in *S.agalactiae* <SEQ ID 2815> which encodes the amino acid sequence <SEQ ID 2816>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 928

A DNA sequence (GBSx0984) was identified in *S.agalactiae* <SEQ ID 2817> which encodes the amino acid sequence <SEQ ID 2818>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1585(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9871> which encodes amino acid sequence <SEQ ID 9872> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA91550 GB:Z67739 DNA topoisomerase IV [Streptococcus pneumoniae] (ver 2)
 Identities = 574/649 (88%), Positives = 617/649 (94%), Gaps = 2/649 (0%)

Query: 5   LAKQDITVTNYGDDAIQVLEGLDAVRKRPGMYIGSTDGTGLHHLVWEIVDNAVDEALSGF 64
           ++K++I + NY DDAIQVLEGLDAVRKRPGMYIGSTDG GLHHLVWEIVDNAVDEALSGF
Sbjct: 1   MSKKEININNYNDDAIQVLEGLDAVRKRPGMYIGSTDGAGLHHLVWEIVDNAVDEALSGF 60

Query: 65  GNRIDVIINKDGSITVTDHGRGMPTGMHAMGKPTVEVIFTVLHAGGKFGQGGYKTSGGLH 124
           G+RIDV INKDGS+TV DHGRGMPTGMHAMG PTVEVIFT+LHAGGKFGQGGYKTSGGLH
Sbjct: 61  GDRIDVTINKDGSLTVQDHGRGMPTGMHAMGIPTVEVIFTILHAGGKFGQGGYKTSGGLH 120

Query: 125 GVGSSVVNALSSWLEVEIIRDGAIYRQRFENGGKPVTTLKKIGTAPKSKSGTSVSFMPDQ 184
           GVGSSVVNALSSWLEVEI RDGA+Y+QRFENGGKPVTTLKKIGTAPKSK+GT V+FMPD
Sbjct: 121 GVGSSVVNALSSWLEVEITRDGAVYKQRFENGGKPVTTLKKIGTAPKSKTGTKVTFMPDA 180

Query: 185 SVFSTIDFKFNTIAERLKESAFLLKNVTLTLTDNRSEEAEHLEFHYENGVQDFVEYLNED 244
           ++FST DFK+NTI+ERL ESAFLLKNVTL+LTD R++EA  +EFHYENGVQDFV YLNED
```

```
Sbjct: 181 TIFSTTDFKYNTISERLNESAFLLKNVTLSLTDKRTDEA--IEFHYENGVQDFVSYLNED 238

Query: 245 KETLTPIMFFEGEEQEFHIEVALQYNDGFSDNILSFVNNVRTKDGGTHETGLKSAITKSM 304
           KE LTP+++FEGE+  F +EVALQYNDGFSDNILSFVNNVRTKDGGTHETGLKSAITK M
Sbjct: 239 KEILTPVLYFEGEDNGFQVEVALQYNDGFSDNILSFVNNVRTKDGGTHETGLKSAITKVM 298

Query: 305 NDYARKTGLLKEKDKNLEGSDYREGLSAILSILVPEEHLQFEGQTKDKLGSPLARPIVDG 364
           NDYARKTGLLKEKDKNLEGSDYREGL+A+LSILVPEEHLQFEGQTKDKLGSPLARP+VDG
Sbjct: 299 NDYARKTGLLKEKDKNLEGSDYREGLAAVLSILVPEEHLQFEGQTKDKLGSPLARPVVDG 358

Query: 365 IVSEKLTYFLMENGDLASNLIRKAIKARDAREAARKARDESRNGKKSKKDKGLLSGKLTP 424
           IV++KLT+FLMENG+LASNLIRKAIKARDAREAARKARDESRNGKK+KKDKGLLSGKLTP
Sbjct: 359 IVADKLTFFLMENGELASNLIRKAIKARDAREAARKARDESRNGKKNKKDKGLLSGKLTP 418

Query: 425 AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNTAKAKMADIIKNEEINT 484
           AQSKN  KNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKV+NTAKAKMADI+KNEEINT
Sbjct: 419 AQSKNPAKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVINTAKAKMADILKNEEINT 478

Query: 485 MIHTIGAGVGPDFNLDDINYDKIIIMTDADTDGAHIQTLLLTFFYRYMRPLVEEGHVYIA 544
           MI+TIGAGVG DF+++D NYDKIIIMTDADTDGAHIQTLLLTFFYRYMRPLVE GHVYIA
Sbjct: 479 MIYTIGAGVGADFSIEDANYDKIIIMTDADTDGAHIQTLLLTFFYRYMRPLVEAGHVYIA 538

Query: 545 LPPLYKMSKGKGKKEIVEYAWTDIELEELRQKFGKGSLLQRYKGLGEMNADQLWETTMNP 604
           LPPLYKMSKGKGKKE V YAWTD ELEELR++FGKG+ LQRYKGLGEMNADQLWETTMNP
Sbjct: 539 LPPLYKMSKGKGKKEEVAYAWTDGELEELRKQFGKGATLQRYKGLGEMNADQLWETTMNP 598

Query: 605 ETRTLIRVTIEDLARAERRVNVLMGDKVPPRRQWIEDNVKFTLEENTVF 653
           ETRTLIRVTIEDLARAERRVNVLMGDKV PRR+WIEDNVKFTLEE TVF
Sbjct: 599 ETRTLIRVTIEDLARAERRVNVLMGDKVEPRRKWIEDNVKFTLEEATVF 647
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2819> which encodes the amino acid sequence <SEQ ID 2820>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1518(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 560/649 (86%), Positives = 615/649 (94%)

Query: 5   LAKQDITVTNYGDDAIQVLEGLDAVRKRPGMYIGSTDGTGLHHLVWEIVDNAVDEALSGF 64
           L K++IT+ NY DDAIQVLEGLDAVRKRPGMYIGSTD TGLHHL+WEIVDNAVDEALSGF
Sbjct: 2   LTKKEITINNYNDDAIQVLEGLDAVRKRPGMYIGSTDATGLHHLIWEIVDNAVDEALSGF 61

Query: 65  GNRIDVIINKDGSITVTDHGRGMPTGMHAMGKPTVEVIFTVLHAGGKFGQGGYKTSGGLH 124
           G+ I V+INKDGS++V D GRGMPTG HAMG PTV+VIFT+LHAGGKFGQGGYKTSGGLH
Sbjct: 62  GDDIKVVINKDGSVSVADSGRGMPTGQHAMGIPTVQVIFTILHAGGKFGQGGYKTSGGLH 121

Query: 125 GVGSSVVNALSSWLEVEIIRDGAIYRQRFENGGKPVTTLKKIGTAPKSKSGTSVSFMPDQ 184
           GVGSSVVNALS+WLEVEI RDG++YRQRFENGGKPVTTLKK+GTAPKSKSGT V+FMPD
Sbjct: 122 GVGSSVVNALSAWLEVEITRDGSVYRQRFENGGKPVTTLKKVGTAPKSKSGTVVTFMPDD 181

Query: 185 SVFSTIDFKFNTIAERLKESAFLLKNVTLTLTDNRSEEAEHLEFHYENGVQDFVEYLNED 244
            +FSTIDFKFNTI+ERLKESAFLLKNV ++LTD R ++    EFHYENGVQDFVEYLNED
Sbjct: 182 KIFSTIDFKFNTISERLKESAFLLKNVKMSLTDLRGDDPIIEEFHYENGVQDFVEYLNED 241

Query: 245 KETLTPIMFFEGEEQEFHIEVALQYNDGFSDNILSFVNNVRTKDGGTHETGLKSAITKSM 304
           KETLTP+++ EG++Q+F +EVALQYNDGFSDNILSFVNNVRTKDGG+HETGLKSAITK+M
Sbjct: 242 KETLTPVIYMEGQDQDFQVEVALQYNDGFSDNILSFVNNVRTKDGGSHETGLKSAITKAM 301

Query: 305 NDYARKTGLLKEKDKNLEGSDYREGLSAILSILVPEEHLQFEGQTKDKLGSPLARPIVDG 364
           NDYARKT LLKEKDKNLEGSDYREGLSA+LSILVPE+HLQFEGQTKDKLGSPLARPIV+
```

```
Sbjct: 302 NDYARKTNLLKEKDKNLEGSDYREGLSAVLSILVPEQHLQFEGQTKDKLGSPLARPIVES 361

Query: 365 IVSEKLTYFLMENGDLASNLIRKAIKARDAREAARKARDESRNGKKSKKDKGLLSGKLTP 424
            IVSEKLT+FL+ENG++AS+L+RKAIKARDAREAARKARD+SRNGKK+KKDKGLLSGKLTP
Sbjct: 362 IVSEKLTFFLLENGEVASHLVRKAIKARDAREAARKARDDSRNGKKNKKDKGLLSGKLTP 421

Query: 425 AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNTAKAKMADIIKNEEINT 484
            AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNT KAKMADI+KNEEINT
Sbjct: 422 AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNTEKAKMADILKNEEINT 481

Query: 485 MIHTIGAGVGPDFNLDDINYDKIIIMTDADTDGAHIQTLLLTFFYRYMRPLVEEGHVYIA 544
            M++TIGAGVG DFNL+DINYDKIIIMTDADTDGAHIQTLLLTFFYRYMRPLVE GHVYIA
Sbjct: 482 MVYTIGAGVGADFNLEDINYDKIIIMTDADTDGAHIQTLLLTFFYRYMRPLVEAGHVYIA 541

Query: 545 LPPLYKMSKGKGKKEIVEYAWTDIELEELRQKFGKGSLLQRYKGLGEMNADQLWETTMNP 604
            LPPLYKMSKGKGK E + YAWTD ELE+LR++FGKG++LQRYKGLGEMNA+QLWETTM+P
Sbjct: 542 LPPLYKMSKGKGKTEKIAYAWTDGELEDLRREFGKGAILQRYKGLGEMNANQLWETTMDP 601

Query: 605 ETRTLIRVTIEDLARAERRVNVLMGDKVPPRRQWIEDNVKFTLEENTVF 653
            ETRTLIRVTI+DLARAERRV+VLMGDK  PRRQWIEDNVKFTLEENTVF
Sbjct: 602 ETRTLIRVTIDDLARAERRVSVLMGDKAAPRRQWIEDNVKFTLEENTVF 650
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 929

A DNA sequence (GBSx0985) was identified in *S.agalactiae* <SEQ ID 2821> which encodes the amino acid sequence <SEQ ID 2822>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80    Transmembrane  378 - 394 ( 378 - 394)

----- Final Results -----
               bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD34369 GB:AF129764 ParC [Streptococcus mitis]
 Identities = 640/820 (78%), Positives = 722/820 (88%), Gaps = 5/820 (0%)

Query: 1   MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPVQRRILYSMNKDGNTFEKGFR 60
            MSNIQNMSLEDIMGERFGRYSKYIIQ+RALPDIRDGLKPVQRRILYSMNKDGNTF+K +R
Sbjct: 1   MSNIQNMSLEDIMGERFGRYSKYIIQDRALPDIRDGLKPVQRRILYSMNKDGNTFDKSYR 60

Query: 61  KSAKSVGNVMGNFHPHGDSSIYDAMVRMSQDWKNRETLIEMHGNNGSMDGDPAAAMRYTE 120
            KSAKSVGN+MGNFHPHGDSSIYDAMVRMSQDWKNRE L+EMHGNNGSMDGDP AAMRYTE
Sbjct: 61  KSAKSVGNIMGNFHPHGDSSIYDAMVRMSQDWKNREILVEMHGNNGSMDGDPPAAMRYTE 120

Query: 121 ARLSEIAGYLLQDIDKNTVPFAWNFDDTEKEPTVLPAAFPNLLVNGATGISAGYATDIPP 180
            ARLSEIAGYLLQDIDK TVPF+WNFDDTEKEPTVLPAAFPNLLVNG+TGISAGYATDIPP
Sbjct: 121 ARLSEIAGYLLQDIDKKTVPFSWNFDDTEKEPTVLPAAFPNLLVNGSTGISAGYATDIPP 180

Query: 181 HNLAEVIDAVVYMIDHPKAKLDKLMEFLPGPDFPTGAIIQGKDEIRKAYETGKGRVAVRS 240
            HNLAEVIDA VYMIDHP AK+DKLMEFLPGPDFPTG IIQG+DEI+KAYETGKGRV VRS
Sbjct: 181 HNLAEVIDAAVYMIDHPTAKVDKLMEFLPGPDFPTGGIIQGRDEIKKAYETGKGRVVVRS 240

Query: 241 RTAIETLKGGKKQIIVTEIPYEVNKSVLVKRIDDVRVNNKVPGIAEVRDESDRDGLRIAI 300
            +T IE LKGGK+QI++TEIPYE+NK+ LVK+IDDVRVN+KV GIAEVRDESDRDGLRIAI
Sbjct: 241 KTEIEKLKGGKEQIVITEIPYEINKANLVKKIDDVRVNSKVAGIAEVRDESDRDGLRIAI 300

Query: 301 ELKKEADETIVLNYLFKYTDLQVNYNFNMVAIDDYTPKQVGLSRILTSYIAHRREIIIAR 360
            ELKK+A+  +VLNYLFKYTDLQ+NYNFNMVAID++TP+QVG+  IL+SYIAHRRE+I+AR
```

```
Sbjct: 301 ELKKDANTELVLNYLFKYTDLQINYNFNMVAIDNFTPRQVGIVPILSSYIAHRREVILAR 360

Query: 361 SKFDKEKAEKRLHIVEGLIRVLSILDEVIALIRASENKADAKENLKVSYEFSEAQAEAIV 420
           S+FDKEKAEKRLHIVEGLIRV+SILDEVIALIRASENKADAKENLKVSY+F+E QAEAIV
Sbjct: 361 SRFDKEKAEKRLHIVEGLIRVISILDEVIALIRASENKADAKENLKVSYDFTEEQAEAIV 420

Query: 421 TLQLYRLTNTDIVTLREEEEELRQQITMLKAIISDERTMYNVMKRELREVKKKFANTRRS 480
           TLQLYRLTNTD+V L+EEE ELR++I ML AII DERTMYN+MK+ELREVKKKFA  R S
Sbjct: 421 TLQLYRLTNTDVVVLQEEEAELREKIAMLAAIIGDERTMYNLMKKELREVKKKFATPRLS 480

Query: 481 ELQELAETIEIDTASLIIEEDTYVSVTRGGYVKRTSPRSFNASTVDELGKREDDELIFVS 540
            L++ A+ IEIDTASLI EEDTYVSVT+ GY+KRTSPRSF AST++E+GKR+DD LIFV
Sbjct: 481 SLEDTAKAIEIDTASLIAEEDTYVSVTKAGYIKRTSPRSFAASTLEEIGKRDDDRLIFVQ 540

Query: 541 NAKTTQHLLMFTNLGNLAYRPVHELADIRWKDVGEHLSQNLVNFASNEEIIYAELVDDF- 599
           +AKTTQHLLMFT LGN+ YRP+HELADIRWKD+GEHLSQ + NF +NEEI+Y E+VD F
Sbjct: 541 SAKTTQHLLMFTTLGNVIYRPIHELADIRWKDIGEHLSQTITNFETNEEILYVEVVDQFD 600

Query: 600 TKETYFAVTSLGQIKRFERQEISPWRTYKSKTAKYAKLKSVEDYVVTVAPIQLEDVILVT 659
              TYFA T LGQIKR ER+E +PWRTYKSK+ KYAKLK   D +V VAPI+L+DV+L++
Sbjct: 601 DATTYFAATRLGQIKRVERKEFTPWRTYKSKSVKYAKLKDDTDQIVAVAPIKLDDVLLIS 660

Query: 660 YNGYALRFSINDVPVVGSKAAGVKAMNLKDRDHIVSAFIANTTSLYLLTHRGSLKRMAID 719
            NGYALRF+I +VPVVG+KAAGVKAMNLK+ D + SAFI NT+S YLLT RGSLKR++ID
Sbjct: 661 QNGYALRFNIEEVPVVGAKAAGVKAMNLKEDDTLQSAFICNTSSFYLLTQRGSLKRVSID 720

Query: 720 VIPTTSRANRGLQVLRELKSKPHRVFKAGPVYLEDSSFEFDLFSSVSNHEGDTFVLEIMS 779
            IP TSRA RGLQVLRELK+KPHRVF AG V   +  F  DLFS+       T  L + S
Sbjct: 721 EIPATSRAKRGLQVLRELKNKPHRVFLAGSV--AEQGFVGDLFSTEVEENDQT--LLVQS 776

Query: 780 KTGKVYDVDLSQWSFSERTSNGSFVSDKISDEEVFSVKIK 819
             G +Y+  L   + SERTSNGSF+SD ISDEEVF   +K
Sbjct: 777 NKGTIYESRLQDLNLSERTSNGSFISDTISDEEVFDAYLK 816
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2823> which encodes the amino acid sequence <SEQ ID 2824>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane  376 - 392 ( 376 - 394)

----- Final Results -----
                 bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 633/819 (77%), Positives = 719/819 (87%)

Query: 1   MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPVQRRILYSMNKDGNTFEKGFR 60
           MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPVQRRILYSMNKDGNTFEKG+R
Sbjct: 3   MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPVQRRILYSMNKDGNTFEKGYR 62

Query: 61  KSAKSVGNVMGNFHPHGDSSIYDAMVRMSQDWKNRETLIEMHGNNGSMDGDPAAAMRYTE 120
           KSAKSVGN+MGNFHPHGDSSIYDAMVRMSQDWKNRE L+EMHGNNGSMDGDP AAMRYTE
Sbjct: 63  KSAKSVGNIMGNFHPHGDSSIYDAMVRMSQDWKNREILVEMHGNNGSMDGDPPAAMRYTE 122

Query: 121 ARLSEIAGYLLQDIDKNTVPFAWNFDDTEKEPTVLPAAFPNLLVNGATGISAGYATDIPP 180
           ARLSEIAGYLLQDI+KNTV FAWNFDDTEKEPTVLPAAFPNLLVNG++GISAGYATDIPP
Sbjct: 123 ARLSEIAGYLLQDIEKNTVSFAWNFDDTEKEPTVLPAAFPNLLVNGSSGISAGYATDIPP 182

Query: 181 HNLAEVIDAVVYMIDHPKAKLDKLMEFLPGPDFPTGAIIQGKDEIRKAYETGKGRVAVRS 240
           HNL+EVIDAVVYMIDHPKA L+KLMEFLPGPDFPTG IIQG DEI+KAYETGKGRV VRS
Sbjct: 183 HNLSEVIDAVVYMIDHPKASLEKLMEFLPGPDFPTGGIIQGADEIKKAYETGKGRVVVRS 242

Query: 241 RTAIETLKGGKKQIIVTEIPYEVNKSVLVKRIDDVRVNNKVPGIAEVRDESDRDGLRIAI 300
```

```
           RT IE LKGGK+QIIVTEIPYEVNK+VLVK+IDDVRVNNKVPGI EVRDESDR GLRIAI
Sbjct: 243 RTEIEELKGGKQQIIVTEIPYEVNKAVLVKKIDDVRVNNKVPGIVEVRDESDRTGLRIAI 302

Query: 301 ELKKEADETIVLNYLFKYTDLQVNYNFNMVAIDDYTPKQVGLSRILTSYIAHRREIIIAR 360
           ELKKEAD   +LNYL KYTDLQVNYNFNMVAID +TP+QVGL +IL+SYI+HR++III R
Sbjct: 303 ELKKEADSQTILNYLLKYTDLQVNYNFNMVAIDHFTPRQVGLQKILSSYISHRKDIIIER 362

Query: 361 SKFDKEKAEKRLHIVEGLIRVLSILDEVIALIRASENKADAKENLKVSYEFSEAQAEAIV 420
           SKFDK KAEKRLHIVEGLIRVLSILDE+IALIR+S+NKADAKENLKVSY+FSE QAEAIV
Sbjct: 363 SKFDKAKAEKRLHIVEGLIRVLSILDEIIALIRSSDNKADAKENLKVSYDFSEEQAEAIV 422

Query: 421 TLQLYRLTNTDIVTLREEEEELRQQITMLKAIISDERTMYNVMKRELREVKKKFANTRRS 480
           TLQLYRLTNTDIVTL+ EE +LR  IT L AII DE TMYNVMKRELREVKKKFAN R S
Sbjct: 423 TLQLYRLTNTDIVTLQNEENDLRDLITTLSAIIGDEATMYNVMKRELREVKKKFANPRLS 482

Query: 481 ELQELAETIEIDTASLIIEEDTYVSVTRGGYVKRTSPRSFNASTVDELGKREDDELIFVS 540
           ELQ  ++ IEIDTASLI EE+T+VSVTRGGY+KRTSPRSFNAS+++E+GKR+DDELIFV
Sbjct: 483 ELQAESQIIEIDTASLIAEEETFVSVTRGGYLKRTSPRSFNASSLEEVGKRDDDELIFVK 542

Query: 541 NAKTTQHLLMFTNLGNLAYRPVHELADIRWKDVGEHLSQNLVNFASNEEIIYAELVDDFT 600
            AKTT+HLL+FT LGN+ YRP+HEL D+RWKD+GEHLSQ + NFA+ EEI+YA++V  F
Sbjct: 543 QAKTTEHLLLFTTLGNVIYRPIHELTDLRWKDIGEHLSQTISNFATEEEILYADIVTSFD 602

Query: 601 KETYFAVTSLGQIKRFERQEISPWRTYKSKTAKYAKLKSVEDYVVTVAPIQLEDVILVTY 660
           +  Y AVT  G IKRF+R+E+SPWRTYKSK+ KY KLK  +D VVT++P+ +ED++LVT
Sbjct: 603 QGLYVAVTQNGFIKRFDRKELSPWRTYKSKSTKYVKLKDDKDRVVTLSPVIMEDLLLVTK 662

Query: 661 NGYALRFSINDVPVVGSKAAGVKAMNLKDRDHIVSAFIANTTSLYLLTHRGSLKRMAIDV 720
           NGYALRFS  +VP+ G K+AGVK +NLK+ D + SAF   + S ++LT RGSLKRMA+D
Sbjct: 663 NGYALRFSSQEVPIQGLKSAGVKGINLKNDDSLASAFAVTSNSFFVLTQRGSLKRMAVDD 722

Query: 721 IPTTSRANRGLQVLRELKSKPHRVFKAGPVYLEDSSFEFDLFSSVSNHEGDTFVLEIMSK 780
           IP TSRANRGL VLRELK+KPHRVF AG V  + S+ +FDLF+ +   E +  +LE++SK
Sbjct: 723 IPQTSRANRGLLVLRELKTKPHRVFLAGGVQSDTSAEQFDLFTDIPEEETNQQMLEVISK 782

Query: 781 TGKVYDVDLSQWSFSERTSNGSFVSDKISDEEVFSVKIK 819
           TG+ Y++ L   S SER SNGSF+SD ISD+EV   + +
Sbjct: 783 TGQTYEIALETLSLSERISNGSFISDTISDQEVLVARTR 821
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 930

A DNA sequence (GBSx0986) was identified in *S.agalactiae* <SEQ ID 2825> which encodes the amino acid sequence <SEQ ID 2826>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3369(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF64593 GB:AF169649 branched-chain aminotransferase IlvE
           [Lactococcus lactis]
 Identities = 259/340 (76%), Positives = 294/340 (86%)

Query: 1   MTVNLDWDNLGFAYRKLPFRYISHFKDGKWDDGKLTDDATLHISESSPALHYGQQAFEGL 60
            M +NLDW+NLGF+YR LPFRYI+ FKDGKW  G+LT D  LHISESSPALHYGQQ FEGL
Sbjct: 1   MAINLDWENLGFSYRNLPFRYIARFKDGKWSAGELTGDNQLHISESSPALHYGQQGFEGL 60

Query: 61  KAYRTKDGSIQLFRPDQNAERLQRTADRLLMPHVPTDKFIAAVKSVVRANEEFVPPYGTG 120
```

```
           KAYRTKDGSIQLFRPDQNA RLQ+TA RL M  V T+ FI AVK VV+AN++FVPPYGTG
Sbjct: 61  KAYRTKDGSIQLFRPDQNAARLQKTARRLCMAEVSTEMFIDAVKQVVKANKDFVPPYGTG 120

Query: 121 ATLYIRPLLIGVGDIIGVKPAEEYIFTVFAMPVGSYFKGGLTPTNFIVSKEYDRAAPNGT 180
           ATLY+RPLLIGVGD+IGVKPA+EYIF VFAMPVGSYFKGGL P+ F++S+EYDRAAP GT
Sbjct: 121 ATLYLRPLLIGVGDVIGVKPADEYIFKVFAMPVGSYFKGGLAPSKFVISREYDRAAPLGT 180

Query: 181 GAAKVGGNYAASLLPGKYAHEKQFSDVIYLDPATHTKIEEVGAANFFGITKDNQFITPLS 240
           G AKVGGNYAASL     A    ++D IYLDP+THTKIEEVGAANFFGIT DN+FITPLS
Sbjct: 181 GGAKVGGNYAASLQAEVGAKASGYADAIYLDPSTHTKIEEVGAANFFGITADNEFITPLS 240

Query: 241 PSILPSITKYSLLYLAKERFGMEAIEGDVFVDELDKFTEAGACGTAAVISPIGGIQNGDD 300
           PSILPSITKYSLLYLA+ R G++AIEG+V+  +L KF EAGACGTAA+ISPIG I +G+D
Sbjct: 241 PSILPSITKYSLLYLAEHRLGLKAIEGEVYAKDLGKFVEAGACGTAAIISPIGRIDDGED 300

Query: 301 FHVFYSETEVGPATRKLYDELVGIQFGDVEAPEGWIYKVD 340
            ++F+SETEVGP  ++LYDELVGIQFGDVEAPEGWI KVD
Sbjct: 301 SYIFHSETEVGPTVKRLYDELVGIQFGDVEAPEGWIVKVD 340
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2827> which encodes the amino acid sequence <SEQ ID 2828>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1208(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 280/340 (82%), Positives = 308/340 (90%)

Query: 1   MTVNLDWDNLGFAYRKLPFRYISHFKDGKWDDGKLTDDATLHISESSPALHYGQQAFEGL 60
           MT+ +DWDNLGF Y KLPFRYIS++K+G+WD G+LT+DATLHISES+PALHYGQQAFEGL
Sbjct: 16  MTIAIDWDNLGFEYHKLPFRYISYYKNGQWDKGQLTEDATLHISESAPALHYGQQAFEGL 75

Query: 61  KAYRTKDGSIQLFRPDQNAERLQRTADRLLMPHVPTDKFIAAVKSVVRANEEFVPPYGTG 120
           KAYRTKDGSIQLFRPD+NA RLQ TADRLLMP V T++FI A K VV+ANE+FVPPYGTG
Sbjct: 76  KAYRTKDGSIQLFRPDRNAVRLQATADRLLMPQVSTEQFIDAAKQVVKANEDFVPPYGTG 135

Query: 121 ATLYIRPLLIGVGDIIGVKPAEEYIFTVFAMPVGSYFKGGLTPTNFIVSKEYDRAAPNGT 180
           ATLY+RPLLIGVGDIIGVKPAEEYIFT+FAMPVG+YFKGGL PTNFIVS+ +DRAAP GT
Sbjct: 136 ATLYLRPLLIGVGDIIGVKPAEEYIFTIFAMPVGNYFKGGLAPTNFIVSEAFDRAAPYGT 195

Query: 181 GAAKVGGNYAASLLPGKYAHEKQFSDVIYLDPATHTKIEEVGAANFFGITKDNQFITPLS 240
           GAAKVGGNYA SLLPGK A    FSDVIYLDPATHTKIEEVGAANFFGIT +N+F+TPLS
Sbjct: 196 GAAKVGGNYAGSLLPGKAAKSAGFSDVIYLDPATHTKIEEVGAANFFGITANNEFVTPLS 255

Query: 241 PSILPSITKYSLLYLAKERFGMEAIEGDVFVDELDKFTEAGACGTAAVISPIGGIQNGDD 300
           PSILPSITKYSLL LA+ER GM  IEGDV ++ELDKF EAGACGTAAVISPIGGIQ  D+
Sbjct: 256 PSILPSITKYSLLQLAEERLGMTVIEGDVPINELDKFVEAGACGTAAVISPIGGIQYKDN 315

Query: 301 FHVFYSETEVGPATRKLYDELVGIQFGDVEAPEGWIYKVD 340
            HVFYSETEVGP TR+LYDELVGIQFGD+EAPEGWI KVD
Sbjct: 316 LHVFYSETEVGPVTRRLYDELVGIQFGDIEAPEGWIVKVD 355
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 931

A DNA sequence (GBSx0987) was identified in *S.agalactiae* <SEQ ID 2829> which encodes the amino acid sequence <SEQ ID 2830>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3459(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9365> which encodes amino acid sequence <SEQ ID 9366> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10915> which encodes amino acid sequence <SEQ ID 10916> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2831> which encodes the amino acid sequence <SEQ ID 2832>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3043(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 22/36 (61%), Positives = 30/36 (83%)

Query: 4  IVSKKDKKIEIQISDAQVTVNGTKVDGYQLVMEKKL 39
           ++SKKDKKIEIQ+ D +V VN TK+DGYQL + K++
Sbjct: 1  VMSKKDKKIEIQLIDHKVMVNETKIDGYQLQIGKRV 36
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 932

A DNA sequence (GBSx0988) was identified in *S.agalactiae* <SEQ ID 2833> which encodes the amino acid sequence <SEQ ID 2834>. This protein is predicted to be glycyl-tRNA synthetase beta subunit (glyS). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB73488 GB:AL139077 glycyl-tRNA synthetase beta chain
           [Campylobacter jejuni]
 Identities = 33/90 (36%), Positives = 49/90 (53%), Gaps = 2/90 (2%)

Query: 3   RAFNLAEKVTHSVLVDSSLFENNQEKALYQAILSLELTEDMHDNLDKLFALSPIINDFFD 62
           R  N+A K  H V  D SLF    E  LY+A         + + L+ LFAL P I++FF+
Sbjct: 570 RLANIATKNPHKV--DESLFVQEAESKLYKAFQEKTKANSLQEKLENLFALKPFIDEFFN 627

Query: 63  NTMVMTDDEKMKQNRLAILNSLVAKARTVA 92
             M+  +DEK+K NR A++  + A+   +A
```

```
Sbjct: 628 QVMINAEDEKLKNNRQALVYEIYAEFLKIA 657
```

There is also homology to SEQ ID 2836.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 933

A DNA sequence (GBSx0989) was identified in *S.agalactiae* <SEQ ID 2837> which encodes the amino acid sequence <SEQ ID 2838>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4825(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13672 GB:Z99113 ynzC [Bacillus subtilis]
 Identities = 41/72 (56%), Positives = 56/72 (76%)

Query: 5  KIARINELSKKKKTVGLTGEEKVEQAKLREEYIEGFRRSVRHHVEGIKLVDDEGNDVTPE 64
           KIARINEL+ K K   +T EEK EQ KLR+EY++GFR S+++ ++ +K++D EGNDVTPE
Sbjct: 6  KIARINELAAKAKAGVITEEEKAEQQKLRQEYLKGFRSSMKNTLKSVKIIDPEGNDVTPE 65

Query: 65 KLRQVQREKGLH 76
           KL++ QR   LH
Sbjct: 66 KLKREQRNNKLH 77
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2839> which encodes the amino acid sequence <SEQ ID 2840>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4303(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 79/85 (92%), Positives = 83/85 (96%)

Query: 1  MDPKKIARINELSKKKKTVGLTGEEKVEQAKLREEYIEGFRRSVRHHVEGIKLVDDEGND 60
           MDPKKIARINEL+KKKKTVGLTG EKVEQAKLREEYIEG+RRSVRHH+EGIKLVD+EGND
Sbjct: 1  MDPKKIARINELAKKKKTVGLTGPEKVEQAKLREEYIEGYRRSVRHHIEGIKLVDEEGND 60

Query: 61 VTPEKLRQVQREKGLHGRSLDDPNS 85
           VTPEKLRQVQREKGLHGRSLDDP S
Sbjct: 61 VTPEKLRQVQREKGLHGRSLDDPKS 85
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 934

A DNA sequence (GBSx0990) was identified in *S.agalactiae* <SEQ ID 2841> which encodes the amino acid sequence <SEQ ID 2842>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2343(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB69985 GB:U94355 glycerol kinase [Enterococcus casseliflavus]
 Identities = 381/496 (76%), Positives = 439/496 (87%)

Query: 3   SEEKYIMAIDQGTTSSRAIIFNKKGEKIASSQKEFPQIFPQAGWVEHNANQIWNSVQSVI 62
           +E+ Y+MAIDQGTTSSRAIIF++ G+KI SSQKEFPQ FP++GWVEHNAN+IWNSVQSVI
Sbjct: 2   AEKNYVMAIDQGTTSSRAIIFDRNGKKIGSSQKEFPQYFPKSGWVEHNANEIWNSVQSVI 61

Query: 63  AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
           AGAFIES I+P  I  IGITNQRETTVVWDK TG PI NAIVWQSRQ++PIADQLK +GH
Sbjct: 62  AGAFIESGIRPEAIAGIGITNQRETTVVWDKTTGQPIANAIVWQSRQSSPIADQLKVDGH 121

Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGLVHV 182
           T MIHEKTGLVIDAYFSATKVRW+LD++ GAQE+A+ GELLFGTID+WLVWKLTDG VHV
Sbjct: 122 TEMIHEKTGLVIDAYFSATKVRWLLDNIEGAQEKADNGELLFGTIDSWLVWKLTDGQVHV 181

Query: 183 TDYSNAARTMLYNIKELKWDDEILELLNIPKAMLPEVKSNSEVYGKTTPFHFYGGEVPIS 242
           TDYSNA+RTMLYNI +L+WD EIL+LLNIP +MLPEVKSNSEVYG T  +HFYG EVPI+
Sbjct: 182 TDYSNASRTMLYNIHKLEWDQEILDLLNIPSSMLPEVKSNSEVYGHTRSYHFYGSEVPIA 241

Query: 243 GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTGEEMQLSQNNLLTTIGYGINGKVHY 302
           GMAGDQQAALFGQ+AFE GM+KNTYGTG+FI+MNTGEE QLS N+LLTTIGYGINGKV+Y
Sbjct: 242 GMAGDQQAALFGQMAFEKGMIKNTYGTGAFIVMNTGEEPQLSDNDLLTTIGYGINGKVYY 301

Query: 303 ALEGSIFIAGSAIQWLRDGLRMIETSSESEGLAQSSTSDDEVYVVPAFTGLGAPYWDSNA 362
           ALEGSIF+AGSAIQWLRDGLRMIETS +SE LA  +  D+EVYVVPAFTGLGAPYWDS A
Sbjct: 302 ALEGSIFVAGSAIQWLRDGLRMIETSPQSEELAAKAKGDNEVYVVPAFTGLGAPYWDSEA 361

Query: 363 RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNNLLMQ 422
           RG+VFGLTRGT+KEDFV+ATLQ++AYQ +DVIDTM+ DSGIDI  L+VDGGAA N+LLMQ
Sbjct: 362 RGAVFGLTRGTTKEDFVRATLQAVAYQSKDVIDTMKKDSGIDIPLLKVDGGAAKNDLLMQ 421

Query: 423 FQADILGIDIARAKNLETTALGAAFLAGLSVGYWESMDELKELNATGQLFQATMNESRKE 482
           FQADIL ID+ RA NLETTALGAA+LAGL+VG+W+ +DELK +   GQ+F   M    ++
Sbjct: 422 FQADILDIDVQRAANLETTALGAAYLAGLAVGFWKDLDELKSMAEEGQMFTPEMPAEERD 481

Query: 483 KLYKGWRKAVKATQVF 498
            LY+GW++AV ATQ F
Sbjct: 482 NLYEGWKQAVAATQTF 497
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2843> which encodes the amino acid sequence <SEQ ID 2844>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2282(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 464/500 (92%), Positives = 484/500 (96%)

Query: 3   SEEKYIMAIDQGTTSSRAIIFNKKGEKIASSQKEFPQIFPQAGWVEHNANQIWNSVQSVI 62
            S+EKYIMAIDQGTTSSRAIIFN+KGEK++SSQKEFPQIFP AGWVEHNANQIWNSVQSVI
Sbjct: 2   SQEKYIMAIDQGTTSSRAIIFNQKGEKVSSSQKEFPQIFPHAGWVEHNANQIWNSVQSVI 61

Query: 63  AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
            AGAFIESSIKP QIEAIGITNQRETTVVWDKKTG+PIYNAIVWQSRQTAPIA+QLKQ+GH
Sbjct: 62  AGAFIESSIKPSQIEAIGITNQRETTVVWDKKTGVPIYNAIVWQSRQTAPIAEQLKQDGH 121

Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGLVHV 182
            T MIHEKTGLVIDAYFSATK+RWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDG VHV
Sbjct: 122 TKMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGAVHV 181

Query: 183 TDYSNAARTMLYNIKELKWDDEILELLNIPKAMLPEVKSNSEVYGKTTPFHFYGGEVPIS 242
            TDYSNAARTMLYNIK+L WDDEILELLNIPK MLPEVKSNSE+YGKT  FHFYGGEVPIS
Sbjct: 182 TDYSNAARTMLYNIKDLTWDDEILELLNIPKDMLPEVKSNSEIYGKTAAFHFYGGEVPIS 241

Query: 243 GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTGEEMQLSQNNLLTTIGYGINGKVHY 302
            GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTG+EMQLS NNLLTTIGYGINGKVHY
Sbjct: 242 GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTGDEMQLSSNNLLTTIGYGINGKVHY 301

Query: 303 ALEGSIFIAGSAIQWLRDGLRMIETSSESEGLAQSSTSDDEVYVVPAFTGLGAPYWDSNA 362
            ALEGSIFIAGSAIQWLRDGL+MIETS ESE  A +STSDDEVYVVPAFTGLGAPYWDSNA
Sbjct: 302 ALEGSIFIAGSAIQWLRDGLKMIETSPESEQFALASTSDDEVYVVPAFTGLGAPYWDSNA 361

Query: 363 RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNNLLMQ 422
            RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNN+LMQ
Sbjct: 362 RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNNMLMQ 421

Query: 423 FQADILGIDIARAKNLETTALGAAFLAGLSVGYWESMDELKELNATGQLFQATMNESRKE 482
            FQADILGIDIARAKNLETTALGAAFLAGL+VGYWE MD LKELNATGQLF+A+MNESRKE
Sbjct: 422 FQADILGIDIARAKNLETTALGAAFLAGLAVGYWEDMDALKELNATGQLFKASMNESRKE 481

Query: 483 KLYKGWRKAVKATQVFAQED 502
            KLYKGW++AVKATQVF QE+
Sbjct: 482 KLYKGWKRAVKATQVFTQEE 501
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 935

A DNA sequence (GBSx0992) was identified in *S.agalactiae* <SEQ ID 2845> which encodes the amino acid sequence <SEQ ID 2846>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3146(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 936

A DNA sequence (GBSx0993) was identified in *S.agalactiae* <SEQ ID 2847> which encodes the amino acid sequence <SEQ ID 2848>. This protein is predicted to be alpha-glycerophosphate oxidase (glpD). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81   Transmembrane   20 -  36 (  20 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC34740 GB:U94770 alpha-glycerophosphate oxidase [Streptococcus pneumoniae]
 Identities = 464/608 (76%), Positives = 539/608 (88%)

Query: 1   MEFSRETRRLALQRMQDRTLDLLIIGGGITGAGVALQAAASGLDTGLIEMQDFAEGTSSR 60
            MEFS++TR L++++MQ+RTLDLLIIGGGITGAGVALQAAASGL+TGLIEMQDFAEGTSSR
Sbjct: 1   MEFSKKTRELSIKKMQERTLDLLIIGGGITGAGVALQAAASGLETGLIEMQDFAEGTSSR 60

Query: 61  STKLVHGGLRYLKQFDVEVVSDTVSERAVVQQIAPHIPKPDPMLLPVYDEPGSTFSMFRL 120
            STKLVHGGLRYLKQFDVEVVSDTVSERAVVQQIAPHIPKPDPMLLPVYDE G+TFS+FRL
Sbjct: 61  STKLVHGGLRYLKQFDVEVVSDTVSERAVVQQIAPHIPKPDPMLLPVYDEDGATFSLFRL 120

Query: 121 KVAMDLYDLLAGVTNTPAANKVLSAEDVLKREPDLQKEGLLGGGVYLDFRNNDARLVIEN 180
            KVAMDLYDLLAGV+NTP ANKVLS + VL+R+P+L+KEGL+GGGVYLDFRNNDARLVIEN
Sbjct: 121 KVAMDLYDLLAGVSNTPTANKVLSKDQVLERQPNLKKEGLVGGGVYLDFRNNDARLVIEN 180

Query: 181 IKRANRDGAYIASHVKAEDFLFDDNNQIIGVRARDLLTDQVIDIKARLVINTTGPWSDTV 240
            IKRAN+DGA IA+HVKAE FLFD++ +I GV ARDLLTDQV +IKARLVINTTGPWSD V
Sbjct: 181 IKRANQDGALIANHVKAEGFLFDESGKITGVVARDLLTDQVFEIKARLVINTTGPWSDKV 240

Query: 241 RNFSNEGKQIHQLRPTKGVHLVVDRQKLNISQPVYVDTGLNDGRMIFVLPREDKTYFGTT 300
            RN SN+G Q  Q+RPTKGVHLVVD  K+ +SQPVY DTGL DGRM+FVLPRE+KTYFGTT
Sbjct: 241 RNLSNKGTQFSQMRPTKGVHLVVDSSKIKVSQPVYFDTGLGDGRMVFVLPRENKTYFGTT 300

Query: 301 DTDYHGDLEHPTVTKEDVDYLLNIVNKRFPEAELTIDDIESSWAGLRPLLSGNSASDYNG 360
            DTDY GDLEHP VT+EDVDYLL IVN RFPE+ +TIDDIESSWAGLRPL++GNSASDYNG
Sbjct: 301 DTDYTGDLEHPKVTQEDVDYLLGIVNNRFPESNITIDDIESSWAGLRPLIAGNSASDYNG 360

Query: 361 GNSGKLSDESFEELIDSVKDYIAHKNHREDVEKAISHVESSTSEKELDPSAVSRGSSFER 420
            GN+G +SDESF+ LI +V+ Y++ +  REDVE A+S +ESSTSEK LDPSAVSRGSS +R
Sbjct: 361 GNNGTISDESFDNLIATVESYLSKEKTREDVESAVSKLESSTSEKHLDPSAVSRGSSLDR 420

Query: 421 DDNGLLTLAGGKITDYRKMAEGAMETIINILDKEYNRKFKLINSKTYPVSGGEINPSNVD 480
            DDNGLLTLAGGKITDYRKMAEGAME +++IL  E++R FKLINSKTYPVSGGE+NP+NVD
Sbjct: 421 DDNGLLTLAGGKITDYRKMAEGAMERVVDILKAEFDRSFKLINSKTYPVSGGELNPANVD 480

Query: 481 SEIEAYAQLGTLSGLSIEDARYIANLYGSNAPKLFALTRQITEAEGLSLVETLSLHYAMD 540
            SEIEA+AQLG   GL  ++A Y+ANLYGSNAPK+FAL   + +A GLSL +TLSLHYAM
Sbjct: 481 SEIEAFAQLGVSRGLDSKEAHYLANLYGSNAPKVFALAHSLEQAPGLSLADTLSLHYAMR 540

Query: 541 YEMALSPTDFFLRRTNHMLFMRDNLDSLIQPVIDEMAKHYQWSDQDKTFYEEELHETLKD 600
             E+ALSP DF LRRTNHMLFMRD+LDS+++PV+DEM + Y W++++K  Y  ++   L +
Sbjct: 541 NELALSPVDFLLRRTNHMLFMRDSLDSIVEPVLDEMGRFYDWTEEEKATYRADVEAALAN 600

Query: 601 NDLAALKD 608
            NDLA LK+
Sbjct: 601 NDLAELKN 608
```

There is also homology to SEQ ID 128.

SEQ ID 2848 (GBS93) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 7; MW 70.6kDa).

GBS93-His was purified as shown in Figure 192, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 937

A DNA sequence (GBSx0994) was identified in *S.agalactiae* <SEQ ID 2849> which encodes the amino acid sequence <SEQ ID 2850>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 938

A DNA sequence (GBSx0995) was identified in *S.agalactiae* <SEQ ID 2851> which encodes the amino acid sequence <SEQ ID 2852>. This protein is predicted to be glycerol uptake facilitator protein (glpF). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.43   Transmembrane  220 - 236 ( 216 - 236)
   INTEGRAL    Likelihood = -6.48   Transmembrane  139 - 155 ( 136 - 158)
   INTEGRAL    Likelihood = -3.88   Transmembrane   87 - 103 (  83 - 107)
   INTEGRAL    Likelihood = -3.03   Transmembrane  164 - 180 ( 162 - 183)

----- Final Results -----
               bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8689> which encodes amino acid sequence <SEQ ID 8690> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
SRCFLG: 0
McG: Length  of  UR:    21
     Peak Value of UR:   2.51
     Net Charge of CR: -2
McG: Discrim Score:      4.43
GvH: Signal Score (-7.5): -0.139999
     Possible site: 50
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 51
ALOM program   count: 4 value:  -7.43 threshold:  0.0
   INTEGRAL    Likelihood = -7.43   Transmembrane  215 - 231 ( 211 - 231)
```

```
   INTEGRAL    Likelihood = -6.48   Transmembrane  134 - 150 ( 131 - 153)
   INTEGRAL    Likelihood = -3.88   Transmembrane   82 -  98 (  78 - 102)
   INTEGRAL    Likelihood = -3.03   Transmembrane  159 - 175 ( 157 - 178)
   PERIPHERAL  Likelihood =  4.98      65
 modified ALOM score:   1.99
icml HYPID: 7  CFP: 0.397

*** Reasoning Step: 3

----- Final Results -----
                 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA91618 GB:U12567 glycerol uptake facilitator [Streptococcus pneumoniae]
 Identities = 150/230 (65%), Positives = 194/230 (84%), Gaps = 1/230 (0%)

Query: 7   DIFGEFLGTALLVLLGNGVVAGVVLPKTKNHNSGWIVITFGWGLAVAIAALVSGNISPAH 66
            ++FGEFLGT +L+LLGNGVVAGVVLPKTK+++SGWIVIT   G+AVA+A  VSG +SPAH
Sbjct: 4   ELFGEFLGTLILILLGNGVVAGVVLPKTKSNSSGWIVITMV-GIAVAVAVFVSGKLSPAH 62

Query: 67  LNPAVSLAFAIKGDLAWGTAILYMIAQIIGAMLGSLLVYLQFRPHYEAAENRADILGTFA 126
            LNPAV++  A+KG L W + + Y++AQ  GAMLG +LV+LQF+PHYEA EN  +IL TF+
Sbjct: 63  LNPAVTIGVALKGGLPWASVLPYILAQFAGAMLGQILVWLQFKPHYEAEENAGNILATFS 122

Query: 127 TGPALKDNFSNFLSEVLGTLVLVLTIFAIGKYNMPPGVGTMSVGMLVVGIGLSLGGTTGY 186
            TGPA+KD  SN +SE+LGT VLVLTIFA+G Y+   G+GT +VG L+VGIGLSLGGTTGY
Sbjct: 123 TGPAIKDTVSNLISEILGTFVLVLTIFALGLYDFQAGIGTFAVGTLIVGIGLSLGGTTGY 182

Query: 187 AINPARDFGPRLLHALLPMKNKGDSDWTYSWIPIVGPMVGAILAALIFAM 236
            A+NPARD GPR++H++LP+ NKGD DW+Y+WIP+VGP++GA LA L+F++
Sbjct: 183 ALNPARDLGPRIMHSILPIPNKGDGDWSYAWIPVVGPVIGAALAVLVFSL 232
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2853> which encodes the amino acid sequence <SEQ ID 2854>. Analysis of this protein sequence reveals the following:

```
          Possible site: 50
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.13   Transmembrane  213 - 229 ( 209 - 232)
   INTEGRAL    Likelihood = -5.52   Transmembrane  137 - 153 ( 132 - 157)
   INTEGRAL    Likelihood = -4.35   Transmembrane  159 - 175 ( 155 - 178)
   INTEGRAL    Likelihood = -1.17   Transmembrane   85 - 101 (  85 - 101)

----- Final Results -----
                 bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA91618 GB:U12567 glycerol uptake facilitator [Streptococcus pneumoniae]
 Identities = 159/230 (69%), Positives = 196/230 (85%), Gaps = 1/230 (0%)

Query: 2   DIFGEFLGTALLVLLGNGVVAGVVLPKTKTHASGWIVIATGWGIAVAVAVFISGKVAPAH 61
            ++FGEFLGT +L+LLGNGVVAGVVLPKTK+++SGWIVI T  GIAVAVAVF+SGK++PAH
Sbjct: 4   ELFGEFLGTLILILLGNGVVAGVVLPKTKSNSSGWIVI-TMVGIAVAVAVFVSGKLSPAH 62

Query: 62  LNPAVSLAFAMSGTIAWSTAIAYSLAQLLGAMVGSTLVFLQFRPHYLAAESQADILGTFA 121
            LNPAV++  A+ G + W++ + Y LAQ  GAM+G  LV+LQF+PHY A E+  +IL TF+
Sbjct: 63  LNPAVTIGVALKGGLPWASVLPYILAQFAGAMLGQILVWLQFKPHYEAEENAGNILATFS 122

Query: 122 TGPAIRDTSSNLLSEIFGTFVLMLGILAFGLYDMPAGLGTLCVGTLVIGIGLSLGGTTGY 181
            TGPAI+DT SNL+SEI GTFVL+L I A GLYD  AG+GT  VGTL++GIGLSLGGTTGY
Sbjct: 123 TGPAIKDTVSNLISEILGTFVLVLTIFALGLYDFQAGIGTFAVGTLIVGIGLSLGGTTGY 182
```

```
Query: 182 AINPARDLGPRLVHAILPLNNKGDSDWSYAWIPVVGPIIGAVLAVLLFQV 231
           A+NPARDLGPR++H+ILP+ NKGD DWSYAWIPVVGP+IGA LAVL+F +
Sbjct: 183 ALNPARDLGPRIMHSILPIPNKGDGDWSYAWIPVVGPVIGAALAVLVFSL 232
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 169/232 (72%), Positives = 202/232 (86%)

Query: 6   MDIFGEFLGTALLVLLGNGVVAGVVLPKTKNHNSGWIVITFGWGLAVAIAALVSGNISPA 65
           MDIFGEFLGTALLVLLGNGVVAGVVLPKTK H SGWIVI  GWG+AVA+A  +SG ++PA
Sbjct: 1   MDIFGEFLGTALLVLLGNGVVAGVVLPKTKTHASGWIVIATGWGIAVAVAVFISGKVAPA 60

Query: 66  HLNPAVSLAFAIKGDLAWGTAILYMIAQIIGAMLGSLLVYLQFRPHYEAAENRADILGTF 125
           HLNPAVSLAFA+ G +AW TAI Y +AQ++GAM+GS LV+LQFRPHY AAE++ADILGTF
Sbjct: 61  HLNPAVSLAFAMSGTIAWSTAIAYSLAQLLGAMVGSTLVFLQFRPHYLAAESQADILGTF 120

Query: 126 ATGPALKDNFSNFLSEVLGTLVLVLTIFAIGKYNMPPGVGTMSVGMLVVGIGLSLGGTTG 185
           ATGPA++D  SN LSE+ GT VL+L I A G Y+MP G+GT+ VG LV+GIGLSLGGTTG
Sbjct: 121 ATGPAIRDTSSNLLSEIFGTFVLMLGILAFGLYDMPAGLGTLCVGTLVIGIGLSLGGTTG 180

Query: 186 YAINPARDFGPRLLHALLPMKNKGDSDWTYSWIPIVGPMVGAILAALIFAMM 237
           YAINPARD GPRL+HA+LP+ NKGDSDW+Y+WIP+VGP++GA+LA L+F +M
Sbjct: 181 YAINPARDLGPRLVHAILPLNNKGDSDWSYAWIPVVGPIIGAVLAVLLFQVM 232
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 939

A DNA sequence (GBSx0996) was identified in *S.agalactiae* <SEQ ID 2855> which encodes the amino acid sequence <SEQ ID 2856>. This protein is predicted to be NADH oxidase. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.87   Transmembrane  152 - 168 ( 152 - 168)

----- Final Results -----
               bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9523> which encodes amino acid sequence <SEQ ID 9524> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA48728 GB:X68847 NADH oxidase [Enterococcus faecalis]
 Identities = 105/423 (24%), Positives = 197/423 (45%), Gaps = 15/423 (3%)

Query: 10  IVILGASFAGMTCAQKLRQLNPNWDIVLIDKEIHPDYVPNGLNWYYRHEISGLNQAMWQT 69
           +V++G + AG +  + +   +P  ++ + ++  +  ++  G+  Y    +       +
Sbjct: 3   VVVVGCTHAGTSAVKSILANHPEAEVTVYERNDNISFLSCGIALYVGGVVKNAADLFYSN 62

Query: 70  EEEQRLQNIRCLFGLKVEKINKEDR-----ELMLSDGSSVYYDQLICAMGSQAESTYIDG 124
            EE             VE+IN +D+      L      +V YD+L+   GS      I G
Sbjct: 63  PEELASLGATVKMEHNVEEINVDDKTVTAKNLQTGATETVSYDKLVMTTGSWPIIPPIPG 122

Query: 125 ADAQGVLTTKTYATSQNAKQVLDKSHKVAVVGAGIIGLDIAYSLHESGKAVTLLEAQERP 184
            DA+ +L  K Y+ +    +    + +V VVG G IG+++  +  ESGK VTL++  +R
Sbjct: 123 IDAENILLCKNYSQANVIIEKAKDAKRVVVVGGGYIGIELVEAFVESGKQVTLVDGLDRI 182

Query: 185 DFRHTDPDMSLPLLDAMAESKLHFFQNQKVEKITVTREEKLCLRTLTGDTFTVDAVILAV 244
             ++ D   +  L   + +  ++    + V++     + K+         F  D VI+ V
```

```
Sbjct: 183 LNKYLDKPFTDVLEKELVDRGVNLALGENVQQFVADEQGKVAKVITPSQEFEADMVIMCV 242

Query: 245 NFRPDSRLLTGLVDLSVDNSVVVNDYFQTSDPNIYAIGDLIWSYFKGLNSAYYMPLINQA 304
            FRP++ LL   VD+  + ++ VN+Y QTS+P+I+A GD    ++    +  Y+PL   A
Sbjct: 243 GFRPNTELLKDKVDMLPNGAIEVNEYMQTSNPDIFAAGDSAVVHYNPSQTKNYIPLATNA 302

Query: 305 IRSAQMLAYHLSGHAVPKLKITRATGSKHFGYYRANIGLT---------ELEAGFYEDTV 355
           +R   ++  +L+   +        +G   FG+   + G+T         ++EA  +ED
Sbjct: 303 VRQGMLVGRNLTEQKLAYRGTQGTSGLYLFGWKIGSTGVTKESAKLNGLDVEATVFEDNY 362

Query: 356 SVTYFPKEQYDLRIKLIANQKTGHLLGAQLISKENCLATANQLVQAISCDMTDFDLAFQD 415
              + P  +  L ++L+  + T  ++G QL+SK +   +AN L  A+   MT  DLA  D
Sbjct: 363 RPEFMPTTEKVL-MELVYEKGTQRIVGGQLMSKYDITQSANTLSLAVQNKMTVEDLAISD 421

Query: 416 FIY 418
           F +
Sbjct: 422 FFF 424
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2857> which encodes the amino acid sequence <SEQ ID 2858>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.35    Transmembrane  155 - 171 ( 155 - 173)

----- Final Results -----
              bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 54-56
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA44611 GB:X62755 NADH peroxidase [Enterococcus faecalis]
 Identities = 111/428 (25%), Positives = 202/428 (46%), Gaps = 24/428 (5%)

Query: 10  VIGASFAGLAFVDKYKDLNPDSQIILIDKESCPNYIPNGINQLFRGDIQDLSDAMWGRAC 69
           V+G+S  G   V++  +L+PD++I   +K    +++  G+     G ++D++     R
Sbjct: 5   VLGSSHGGYEAVEELLNLHPDAEIQWYEKGDFISFLSCGMQLYLEGKVKDVNSV---RYM 61

Query: 70  LAAQIESN--HRFIQAEVLAIEAPSNTLLLKDS-QGRVFEEGYETLVCAMGASPQSHYIE 126
              ++ES   + F   E+ AI+   + + +KD   G    E Y+ L+ + GA P    I
Sbjct: 62  TGEKMESRGVNVFSNTEITAIQPKEHQVTVKDLVSGEERVENYDKLIISPGAVPFELDIP 121

Query: 127 TSQTNKVLVTKYYEESQASLKLIEASQE-----VLVIGAGLIGLDLAYSLSLQGKRVKLI 181
               + + + +     Q ++KL + + +     V+VIG+G IG++ A + +  GK+V +I
Sbjct: 122 GKDLDNIYLMR---GRQWAIKLKQKTVDPEVNNVVVIGSGYIGIEAAEAFAKAGKKVTVI 178

Query: 182 EAAERPDFYQTDAELIAPVMAEMSTHHVTFINNKRVTAIHEIEGKVVAHTEQGDTFQGDL 241
           +  +RP     D E    +  EM  +++T    + V   +E +G+V       + +  DL
Sbjct: 179 DILDRPLGVYLDKEFTDVLTEEMEANNITIATGETVER-YEGDGRVQKVVTDKNAYDADL 237

Query: 242 AILAINFRPNTHLLQGQVACALDKTILVNENLQTSQANIYAIGDMVSLHFGILGMDYYTP 301
            ++A+  RPNT  L+G +    +  I  +E ++TS+ +++A+GD   + +     +
Sbjct: 238 VVVAVGVRPNTAWLKGTLELHPNGLIKTDEYMRTSEPDVFAVGDATLIKYNPADTEVNIA 297

Query: 302 LINQAMKTGQALALHLAGYPIPPLQTVK-VLGSSHFDYYRASVGVTE-------EEAELY 353
           L   A K G+    +L   P+ P   V+   G + FDY  AS G+ E       +E +
Sbjct: 298 LATNARKQGRFAVKNLE-EPVKPFPGVQGSSGLAVFDYKFASTGINEVMAQKLGKETKAV 356

Query: 354 MDTCSYLYQNGDSKNLFWLKLIARKTDGILIGAQLLSKTNALVIANQLGQALALKVTDAD 413
                YL      K   W KL+       ++GAQL+SK +     N +  A+  K+T  D
Sbjct: 357 TVVEDYLMDFNPDKQKAWFKLVYDPETTQILGAQLMSKADLTANINAISLAIQAKMTIED 416

Query: 414 LAFQDFLF 421
```

```
           LA+ DF F
Sbjct: 417 LAYADFFF 424
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 192/440 (43%), Positives = 276/440 (62%), Gaps = 7/440 (1%)

Query: 8   KVIVILGASFAGMTCAQKLRQLNPNWDIVLIDKEIHPDYVPNGLNWYYRHEISGLNQAMW 67
            K I ++GASFAG+    K + LNP+  I+LIDKE  P+Y+PNG+N  +R +I  L+ AMW
Sbjct: 6   KTIHVIGASFAGLAFVDKYKDLNPDSQIILIDKESCPNYIPNGINQLFRGDIQDLSDAMW 65

Query: 68  -QTEEEQRLQNIRCLFGLKVEKINKEDRELMLSDGSSVY----YDQLICAMGSQAESTYI 122
             +     ++++       +V  I      L+L D         Y+ L+CAMG+  +S YI
Sbjct: 66  GRACLAAQIESNHRFIQAEVLAIEAPSNTLLLKDSQGRVFEEGYETLVCAMGASPQSHYI 125

Query: 123 DGADAQGVLTTKTYATSQNAKQVLDKSHKVAVVGAGIIGLDIAYSLHESGKAVTLLEAQE 182
            + +    VL TK Y  SQ + ++++ S +V V+GAG+IGLD+AYSL   GK V L+EA E
Sbjct: 126 ETSQTNKVLVTKYYEESQASLKLIEASQEVLVIGAGLIGLDLAYSLSLQGKRVKLIEAAE 185

Query: 183 RPDFRHTDPDMSLPLLDAMAESKLHFFQNQKVEKITVTREEKLCLRTLTGDTFTVDAVIL 242
            RPDF  TD ++  P++  M+   + F  N++V  I    E K+   T  GDTF  D  IL
Sbjct: 186 RPDFYQTDAELIAPVMAEMSTHHVTFINNKRVTAIHEI-EGKVVAHTEQGDTFQGDLAIL 244

Query: 243 AVNFRPDSRLLTGLVDLSVDNSVVVNDYFQTSDPNIYAIGDLIWSYFKGLNSAYYMPLIN 302
            A+NFRP++ LL G V  ++D +++VN+  QTS  NIYAIGD++  +F  L   YY PLIN
Sbjct: 245 AINFRPNTHLLQGQVACALDKTILVNENLQTSQANIYAIGDMVSLHFGILGMDYYTPLIN 304

Query: 303 QAIRSAQMLAYHLSGHAVPKLKITRATGSKHFGYYRANIGLTELEAGFYEDTVSVTYFPK 362
            QA+++ Q LA HL+G+ +P L+  +  GS HF YYRA++G+TE EA  Y DT S  Y
Sbjct: 305 QAMKTGQALALHLAGYPIPPLQTVKVLGSSHFDYYRASVGVTEEEAELYMDTCSYLYQNG 364

Query: 363 EQYDL-RIKLIANQKTGHLLGAQLISKENCLATANQLVQAISCDMTDFDLAFQDFIYTAR 421
            +  +L  +KLIA +  G L+GAQL+SK N L  ANQL QA++  +TD DLAFQDF++
Sbjct: 365 DSKNLFWLKLIARKTDGILIGAQLLSKTNALVIANQLGQALALKVTDADLAFQDFLFLQG 424

Query: 422 ESEMAYMLHQAAINLYEKRI 441
             S++AY LH+A + L+EKR+
Sbjct: 425 HSDLAYHLHEACLKLFEKRL 444
```

There is also homology to SEQ IDs 1820, 1876, 4666.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 940

A DNA sequence (GBSx0998) was identified in *S.agalactiae* <SEQ ID 2859> which encodes the amino acid sequence <SEQ ID 2860>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2980(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 941

A DNA sequence (GBSx0999) was identified in *S.agalactiae* <SEQ ID 2861> which encodes the amino acid sequence <SEQ ID 2862>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3548(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 942

A DNA sequence (GBSx1000) was identified in *S.agalactiae* <SEQ ID 2863> which encodes the amino acid sequence <SEQ ID 2864>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1685(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9525> which encodes amino acid sequence <SEQ ID 9526> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2865> which encodes the amino acid sequence <SEQ ID 2866>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3125(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 179/476 (37%), Positives = 279/476 (58%), Gaps = 5/476 (1%)

Query: 1   MRIEALMEKERRVQYRLLSFLRGSPQAIALKLALLETGLSRATFLKYINNLNSYFEQEKV 60
           M+IE LM+KERR QYRLL  L  + + +  LK  +  + LS+ T LKYI+NLN    ++ +
Sbjct: 21  MKIEDLMDKERRAQYRLLVTLYHAKETLRLKDLMRLSNLSKVTLLKYIDNLNHLCREQGL 80

Query: 61  NCRIVYYKDKLFLEEDYNLSNQEVLKALMKDSIKYTILISLFNQRQFTIVGLSQELMVSE 120
            C+++  KD L L+E+     ++++  L+K+S+ Y IL  ++    F I  LS ELMVSE
Sbjct: 81  ACQLLLEKDSLSLKENGQFHWEDLVALLLKESVAYQILTYMYCHEHFNITNLSVELMVSE 140

Query: 121 ATLNRHLAHLNELLAEFDIAISQGKQIGDELQWRYFYYELFKQLWSYDKCQNMIKKLDLD 180
           ATLNR LAHLN+LL+EFD+A+SQG+Q+G ELQWRYFY+ELF+   +      ++ +LD
```

```
Sbjct: 141 ATLNRQLAHLNQLLSEFDLALSQGRQLGSELQWRYFYFELFRHTLTRQGIDALVNQLDAS 200

Query: 181 SLILLIERLAQHTLTREAHQNLGLWFSICHHRLLAMEKISDNLKPIVKHYQCNAFYKRLD 240
            L  LIERL   +L+ EA + L +W +I   R+   +  +D+          N F+KRL+
Sbjct: 201 HLATLIERLIGQSLSAEALEQLLIWLAISQARMSFQKSYNDHFLRDSDFMTSNIFFKRLE 260

Query: 241 AALVLYMSRFALEYREGEVLATFAFLHSQNILPINTMEYIMGFGGPIIDCVTETIIYFKK 300
           + L+ Y+ R+ALE+   E  + F FLH+  +LPI +M+Y +GFGGPI D ++E +   KK
Sbjct: 261 SMLLHYLRRYALEFDAFEAKSLFVFLHAYPLLPIASMKYSLGFGGPIADHISEALWLLKK 320

Query: 301 ESILADETSDQVIYQLGQLYSHYYFFKGHILVEQPDLEQTYRLIDHNMRDKLHHISKKII 360
             ++  +T +++IY LG  +S  YFFKG IL +  + +  Y+L+  + R  L  I   ++
Sbjct: 321 AHVIIHQTKEEIIYGLGIFFSKAYFFKGAILSQPTNSQYLYQLVGEDKRALLRVIINHLV 380

Query: 361 ANVNRIRPLTEDGCSLLTLHLLELLIFSKNSQKMPFRIGLDMTGNAVEQSLLEYRIRQHF 420
             +++      D    L+  +L LLIFS      P  +GL +  N VE ++ E  IR+H
Sbjct: 381 LQMDQ----ETDFSQQLSDDILALLIFSIERHHEPLLVGLALGQNKVEAAIAELAIRRHL 436

Query: 421 SGNNSIQVEPYDEGKGFD-MVIYQSHSRPYKAKLTYCLNKGASERELQEIDSLIYD 475
                 Q+ PYD  K +D ++ YQ+   P +    Y L + +S  EL  +++ + D
Sbjct: 437 GHRRDFQLMPYDHQKVYDCLITYQTVCLPRQDLPYYRLKQYSSPYELTALEAFLKD 492
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 943

A DNA sequence (GBSx1001) was identified in *S.agalactiae* <SEQ ID 2867> which encodes the amino acid sequence <SEQ ID 2868>. This protein is predicted to be transketolase (tktA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2084(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9527> which encodes amino acid sequence <SEQ ID 9528> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06071 GB:AP001515 transketolase [Bacillus halodurans]
 Identities = 403/661 (60%), Positives = 520/661 (77%), Gaps = 8/661 (1%)

Query: 6   IDQLAVNTVRTLSIDAIQAANSGHPGLPMGAAPMAYVLWNKFLNVNPKTSRNWTNRDRFV 65
            ++QLAVNT+RTLSID+++ ANSGHPG+PMGAAPMA+ LW KF+N NP  + +W NRDRFV
Sbjct: 5   VEQLAVNTIRTLSIDSVEKANSGHPGMPMGAAPMAFCLWTKFMNHNP-ANPDWVNRDRFV 63

Query: 66  LSAGHGSALLYSLLHLAGYDLSIDDLKQFRQWGSKTPGHPEVNHTDGVEATTGPLGQGIA 125
            LSAGHGS LLYSLLHL GYDLS+++L+ FRQWGSKTPGHPE  HT GVEATTGPLGQG+A
Sbjct: 64  LSAGHGSMLLYSLLHLTGYDLSLEELQNFRQWGSKTPGHPEYGHTPGVEATTGPLGQGVA 123

Query: 126 NAVGMAMAEAHLAAKFNKPGFDLVDHYTYTLHGDGCLMEGVSQEAASLAGHLKLGKLVLL 185
             AVGMAMAE HLAA +N+ G+++VDHYTYT+ GDG LMEGVS EAASLAGHLKLG+++LL
Sbjct: 124 MAVGMAMAERHLAATYNRDGYNIVDHYTYTICGDGDLMEGVSAEAASLAGHLKLGRMILL 183

Query: 186 YDSNDISLDGPTSQSFTEDVKGRFESYGWQHILVKDGNDLEAIAAAIEAAKAETDKPTII 245
            YDSNDISLDG    SF+E V+ RF++YGW  + V+DGN+L+ IA AIE AKA+ ++P++I
Sbjct: 184 YDSNDISLDGDLHHSFSESVEDRFKAYGWHVVRVEDGNNLDEIAKAIEEAKAD-ERPSLI 242

Query: 246 EVKTIIGFGAEKQGTSSV-HGAPLGAEGITFAKKAYVWEYP-DFTVPAEVADRFASDLQA 303
```

```
           EVKT IGFG+  +G  SV HGAPLGA+ +   K+AY W Y  +F +P EVA  +   ++
Sbjct: 243 EVKTTIGFGSPNKGGKSVSHGAPLGADEVKLTKEAYEWTYENEFHIPEEVA-AYYEQVKQ 301

Query: 304 RGAKAEEAWNDLFAKYEVEYPELATEYKEAFAG---QAETVELKAHDLGSSVASRVSSQQ 360
           +GA+ EE+WN+LFA+Y+  YPELA++++ A  G   +       ++++G SVA+R SS +
Sbjct: 302 QGAEKEESWNELFAQYKKAYPELASQFELAVHGDLPEGWDAVAPSYEVGKSVATRSSSGE 361

Query: 361 AIQQLSTQLPNLWGGSADLSASNNTMVAAETDFQASNYAGRNIWFGVREFAMAAAMNGIA 420
           A+   +  +P L+GGSADL++SN T++  E +F   +Y+GRN+WFGVREFAM AAMNG+A
Sbjct: 362 ALNAFAKTVPQLFGGSADLASSNKTLIKGEANFSRDDYSGRNVWFGVREFAMGAAMNGMA 421

Query: 421 LHGGTRVYGGTFFVFSNYLLPAVRMAALQNLPTVYVMTHDSIAVGEDGPTHEPIEQLASV 480
           LHGG +V+G TFFVFS+YL PA+R+AAL  LP +YV THDSIAVGEDGPTHEP+EQLAS+
Sbjct: 422 LHGGLKVFGATFFVFSDYLRPAIRLAALMQLPVIYVFTHDSIAVGEDGPTHEPVEQLASL 481

Query: 481 RSMPNLNVIRPADGNETNAAWQRAVSETDRPTMLVLTRQNLPVLEGTSELAQEGVNKGAY 540
           R+MP L+VIRPADGNE+ AAW+ A+   D+PT LVL+RQNLP LEG  + A +GV+KGAY
Sbjct: 482 RAMPGLSVIRPADGNESVAAWKLALESKDQPTALVLSRQNLPTLEGAVDRAYDGVSKGAY 541

Query: 541 ILSEAKGELDGIIIATGSEVKLALDTQDKLESEGIHVRVVSMPAQNIFDEQEASYQEQVL 600
           +L+ A G  D +++A+GSEV LA++ ++ LE EGIH  VVSMP+ + F+ Q A Y+E+VL
Sbjct: 542 VLAPANGSADLLLLASGSEVSLAVNAKEALEKEGIHAAVVSMPSWDRFEAQSAEYKEEVL 601

Query: 601 PSAVTKRLAIEAGSSFGWGKYVGLNGLTLTIDTWGASAPGNRIFEEYGFTVENAVSLYKEL 661
           PS VT RLAIE GSS GW KYVG  G  + ID +GASAPG RI EE+GFTV++ V+  K L
Sbjct: 602 PSDVTARLAIEMGSSLGWAKYVGNQGDVVAIDRFGASAPGERIMEEFGFTVQHVVARAKAL 662
```

There is also homology to SEQ ID 520.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 944

A DNA sequence (GBSx1002) was identified in *S.agalactiae* <SEQ ID 2869> which encodes the amino acid sequence <SEQ ID 2870>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4477(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9529> which encodes amino acid sequence <SEQ ID 9530> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2871> which encodes the amino acid sequence <SEQ ID 2872>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4581(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 27/79 (34%), Positives = 45/79 (56%)
```

```
Query: 3  MKKECRDFYRQIQHTYNDISVREDAVLSSILLSASNGLIKTSDVPRVAYELTQQLENNEI 62
          M+K+ +  Y  I+  Y+    RE+  LS +LL+ASN LIK S+   VAY+L Q ++N  +
Sbjct: 1  MEKKRQRLYDVIRQAYDYPENRENVALSQLLLAASNRLIKHSNPLLVAYQLNQDVDNYLL 60

Query: 63 EKSFESLATVKELKKSAKK 81
          +        ++  K+S +K
Sbjct: 61 DNDILLPKSLCRFKQSLEK 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 945

A DNA sequence (GBSx1003) was identified in *S.agalactiae* <SEQ ID 2873> which encodes the amino acid sequence <SEQ ID 2874>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2610(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB49925 GB:AJ248286 ABC transporter, ATP-binding protein
          [Pyrococcus abyssi]
 Identities = 96/243 (39%), Positives = 164/243 (66%), Gaps = 2/243 (0%)

Query: 1   MIKFEHVSKVYGEKEALSDLTLSVKDGEIFGLIGHNGAGKTTTISILTSIIDATYGQVYI 60
            MI  E++ K +G KE L  ++ +VKDGEI+GL+G NG+GK+TT+ IL+ II    G+V +
Sbjct: 1   MIIVENLRKRFGGKEVLKGISFTVKDGEIYGLLGPNGSGKSTTMRILSGIITDFEGKVIV 60

Query: 61  DDLLLTEHRDQIKKKIGYVPDSPDIFLNLTAEEYWYFLAKIYDVAPEDIEARITKLVDIF 120
              + + +   Q+K+ +GYVP++P ++ +LT  E++ F+  +  +  + +E R+ KLV+ F
Sbjct: 61  GGVEVAKDPLQVKRIVGYVPETPALYESLTPAEFFSFVGGVRGIPKDILEERVRKLVEAF 120

Query: 121 ELEEQRYNPIESFSHGMRQKVIVIGALLPNPDIWILDEPLTGLDPQASFDLKEMMKEHAK 180
            E+++     I + S G +QK+ +I +LL +P + ILDE + GLDP+++   +E++ E  +
Sbjct: 121 EIKKYMNQLIGTLSFGTKQKISLISSLLHDPKVLILDEAMNGLDPKSARIFRELLYEFKE 180

Query: 181 NGKTVIFSTHVLAVAEQLCDRIGILKQGKLIFVGSLGELKMKYPDKDLETIYLELAGRQA 240
             GK+++FSTHVLA+AE +CDR+GI+ QG++I  G++ ELK    ++ LE ++L+L   QA
Sbjct: 181 EGKSIVFSTHVLALAELICDRVGIIYQGRIIAEGTVEELKEISKEERLEDVFLKLT--QA 238

Query: 241 SRE 243
              E
Sbjct: 239 KEE 241
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2875> which encodes the amino acid sequence <SEQ ID 2876>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2723(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 182/244 (74%), Positives = 215/244 (87%)

Query: 1   MIKFEHVSKVYGEKEALSDLTLSVKDGEIFGLIGHNGAGKTTTISILTSIIDATYGQVYI 60
           MI+F+HVSK+YG+KEALSDL +++ DGEIFGLIGHNGAGKTTTISILTSII+A+YG+V++
Sbjct: 1   MIEFKHVSKLYGDKEALSDLNVTINDGEIFGLIGHNGAGKTTTISILTSIIEASYGEVFV 60

Query: 61  DDLLLTEHRDQIKKKIGYVPDSPDIFLNLTAEEYWYFLAKIYDVAPEDIEARITKLVDIF 120
           D  LLTE+R+ IKK+I YVPDSPDIFLNLT  EYW FLAKIY V+ ED E R+ +L  +F
Sbjct: 61  DGQLLTENREAIKKQIAYVPDSPDIFLNLTPNEYWQFLAKIYGVSDEDREERLAQLTTLF 120

Query: 121 ELEEQRYNPIESFSHGMRQKVIVIGALLPNPDIWILDEPLTGLDPQASFDLKEMMKEHAK 180
           EL+E+    I+SFSHGMRQKVIVIGAL+ NP+IWILDEPLTGLDPQASFDLKEMMK HA
Sbjct: 121 ELKEEVNQTIDSFSHGMRQKVIVIGALVSNPNIWILDEPLTGLDPQASFDLKEMMKAHAA 180

Query: 181 NGKTVIFSTHVLAVAEQLCDRIGILKQGKLIFVGSLGELKMKYPDKDLETIYLELAGRQA 240
           +G TV+FSTHVL+VAEQLCDRIGILK+GKLIFVG++ ELK  +PDKDLE+IYLELAGR+A
Sbjct: 181 SGHTVLFSTHVLSVAEQLCDRIGILKKGKLIFVGTIDELKEHHPDKDLESIYLELAGRKA 240

Query: 241 SREG 244
             EG
Sbjct: 241 QEEG 244
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 946

A DNA sequence (GBSx1004) was identified in *S.agalactiae* <SEQ ID 2877> which encodes the amino acid sequence <SEQ ID 2878>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.43   Transmembrane  504 - 520 ( 495 - 529)
   INTEGRAL    Likelihood =-12.58   Transmembrane  427 - 443 ( 400 - 449)
   INTEGRAL    Likelihood =-10.99   Transmembrane  151 - 167 ( 144 - 179)
   INTEGRAL    Likelihood = -8.44   Transmembrane  194 - 210 ( 189 - 214)
   INTEGRAL    Likelihood = -7.96   Transmembrane   48 -  64 (  46 -  68)
   INTEGRAL    Likelihood = -7.32   Transmembrane  350 - 366 ( 348 - 378)
   INTEGRAL    Likelihood = -6.69   Transmembrane  475 - 491 ( 474 - 501)
   INTEGRAL    Likelihood = -6.00   Transmembrane  319 - 335 ( 318 - 337)
   INTEGRAL    Likelihood = -5.73   Transmembrane  252 - 268 ( 244 - 271)
   INTEGRAL    Likelihood = -4.78   Transmembrane  125 - 141 ( 121 - 148)
   INTEGRAL    Likelihood = -4.51   Transmembrane   76 -  92 (  71 -  98)
   INTEGRAL    Likelihood = -3.56   Transmembrane  406 - 422 ( 400 - 426)

------ Final Results -----
              bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2879> which encodes the amino acid sequence <SEQ ID 2880>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.33   Transmembrane  167 - 183 ( 158 - 193)
   INTEGRAL    Likelihood =-12.52   Transmembrane  524 - 540 ( 508 - 546)
   INTEGRAL    Likelihood =-10.93   Transmembrane   63 -  79 (  60 -  84)
   INTEGRAL    Likelihood = -8.39   Transmembrane  421 - 437 ( 414 - 456)
   INTEGRAL    Likelihood = -8.23   Transmembrane  208 - 224 ( 203 - 228)
   INTEGRAL    Likelihood = -8.23   Transmembrane  504 - 520 ( 493 - 521)
   INTEGRAL    Likelihood = -7.59   Transmembrane  139 - 155 ( 134 - 162)
   INTEGRAL    Likelihood = -6.64   Transmembrane  261 - 277 ( 257 - 287)
```

```
    INTEGRAL    Likelihood = -4.99   Transmembrane  446 - 462 ( 444 - 464)
    INTEGRAL    Likelihood = -4.25   Transmembrane  369 - 385 ( 367 - 387)
    INTEGRAL    Likelihood = -0.80   Transmembrane   87 - 103 (  87 - 104)
    INTEGRAL    Likelihood = -0.11   Transmembrane  334 - 350 ( 334 - 350)

 ----- Final Results -----
             bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9173> which encodes the amino acid sequence <SEQ ID 9174>. Analysis of this protein sequence reveals the following:

```
     Possible site: 51
 >>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood =-14.33   Transmembrane  153 - 169 ( 144 - 179)
     INTEGRAL    Likelihood =-12.52   Transmembrane  510 - 526 ( 494 - 532)
     INTEGRAL    Likelihood =-10.93   Transmembrane   49 -  65 (  46 -  70)
     INTEGRAL    Likelihood = -8.39   Transmembrane  407 - 423 ( 400 - 442)
     INTEGRAL    Likelihood = -8.23   Transmembrane  194 - 210 ( 189 - 214)
     INTEGRAL    Likelihood = -8.23   Transmembrane  490 - 506 ( 479 - 507)
     INTEGRAL    Likelihood = -7.59   Transmembrane  125 - 141 ( 120 - 148)
     INTEGRAL    Likelihood = -6.64   Transmembrane  247 - 263 ( 243 - 273)
     INTEGRAL    Likelihood = -4.99   Transmembrane  432 - 448 ( 430 - 450)
     INTEGRAL    Likelihood = -4.25   Transmembrane  355 - 371 ( 353 - 373)
     INTEGRAL    Likelihood = -0.80   Transmembrane   73 -  89 (  73 -  90)
     INTEGRAL    Likelihood = -0.11   Transmembrane  320 - 336 ( 320 - 336)

 ----- Final Results -----
             bacterial membrane --- Certainty=0.673(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 255/542 (47%), Positives = 378/542 (69%), Gaps = 12/542 (2%)

Query: 1   MNWSRIWELVKINILYSNPQTLSALRKKQEKHPKKEFSAYKSMFRNQLFQILLFSIIYVF 60
           MNWS IWEL+KINILYSNPQ+L+ L+K+QEKHPK+ F AYKSM R Q   I +F +IY+F
Sbjct: 15  MNWSTIWELIKINILYSNPQSLANLKKRQEKHPKENFKAYKSMMRQQALMIAMFLVIYLF 74

Query: 61  LFVSLDFKEYPGYFTFYIGIFTLVSIIYSFIAMYSVFYESDDVKQYAYLPIKSEELYVAK 120
           +F+ +DF  YPG F+F + +F ++S + +F ++Y++FYES+D+K Y +LP+ SEELY+AK
Sbjct: 75  MFIGVDFSHYPGLFSFDVAMFFIMSTLTAFSSLYTIFYESNDLKLYIHLPVTSEELYIAK 134

Query: 121 IFATFGMSVTFLMPILTLMIVAYWRIIGGPLAVLLAIINFAILFLSVTVISLYINSLIGR 180
           I ++ GM   FLMP+++L+++AYW+++G PL++L+AI+ F +L +S  V+++YIN+ +G+
Sbjct: 135 IVSSLGMGAVFLMPLISLLLIAYWQLLGNPLSILVAIVLFLVLLVSSMVLAIYINAWVGK 194

Query: 181 AIIRSANRKLISTILISLATFGAIVPLLFVNMTSQK--MVQGKLQDIAPIPYVRGYYDIV 238
            I+RS  RKLISTI++ ++TFGA V +  +N+++ K  M  G   D   IPY +G+YD+V
Sbjct: 195 IIVRSRKRKLISTIMMFVSTFGAFVLIFAINISNNKRTMTDGVFTDYPTIPYFKGFYDVV 254

Query: 239 TAPFSMESLLNYYLPLLIILFLIGAIYKWVMPRYYQELLY----GQVKQRK--VHRQIDF 292
            APFS  +LLN++LPLL+IL ++  I   VMP YY+E  Y     +VKQ K  V+R
Sbjct: 255 QAPFSTAALLNFWLPLLLILAMVYGIVTKVMPTYYREAFYISNENKVKQTKKPVNRP--- 311

Query: 293 SKRESINKTLVKHHLSSLQNATLLTNTFLMPLLYLAMFIVPILNNGKEIGRFFNENYFGI 352
            + +S+ + L KHHL +LQNATLLT T+LMPL+Y+ +FI P L+ G    +  + +YFG+
Sbjct: 312 HQNQSLAQLLRKHHLLTLQNATLLTQTYLMPLMYVMLFIGPSLSRGTGFFKHISPDYFGV 371

Query: 353 AFLAGILIGSLCVMPASIVGVGISLEKSNFYFIKSLPISFSYFLKHKFVTLITLQLAVPT 412
           A L G+ +G +C  P S +GVGISLEK NF FIKSLPI+   FL  KF  L+ LQL VP
Sbjct: 372 ALLFGVSLGVMCATPTSFIGVGISLEKDNFTFIKSLPITLKKFLMDKFCLLVGLQLIVPM 431

Query: 413 FIYFLVGFFLLKLSILVLLSFILGLVFMGLIEGQFIYRRDYKHLFLNWQEVTQLFNRGLG 472
            IY + G F+L L  L+ ++F LG     +++G+ +YRRDY+ L L WQ++TQLF RG G
```

```
Sbjct: 432 VIYLVFGLFVLHLHPLLTIAFCLGYALSLIVQGELMYRRDYRLLDLKWQDMTQLFTRGDG 491

Query: 473 QWLLVGSLFGMMIIGSFL-IGISIFWSMVWNTVAVNIIILIIGLLILSICQYLLLKNFWK 531
           QWL +G +FG +I+   L  G  I  +++   + ++I++  + L++L + Q  + K FWK
Sbjct: 492 QWLTMGLIFGNLIVAGVLGFGAVIIANIIQQPLLISILLSCLILMVLGLAQLWIQKTFWK 551

Query: 532 KL 533
            L
Sbjct: 552 SL 553
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 947

A DNA sequence (GBSx1005) was identified in *S.agalactiae* <SEQ ID 2881> which encodes the amino acid sequence <SEQ ID 2882>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.12    Transmembrane  242 - 258  ( 239 - 265)
   INTEGRAL    Likelihood = -7.64    Transmembrane  430 - 446  ( 421 - 450)
   INTEGRAL    Likelihood = -5.84    Transmembrane  120 - 136  ( 113 - 139)
   INTEGRAL    Likelihood = -5.52    Transmembrane  212 - 228  ( 210 - 232)
   INTEGRAL    Likelihood = -5.20    Transmembrane  287 - 303  ( 283 - 313)
   INTEGRAL    Likelihood = -3.56    Transmembrane  148 - 164  ( 143 - 166)
   INTEGRAL    Likelihood = -0.48    Transmembrane  382 - 398  ( 382 - 398)

----- Final Results -----
            bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15963 GB:Z99124 phosphotransferase system (PTS)
           beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]
 Identities = 175/447 (39%), Positives = 266/447 (59%), Gaps = 10/447 (2%)

Query: 4   EYITLSKNIIKHLGGQNNINNVYHCQTRLRFSLNDPTKVNLEQLKTLKEVKTVVISGGQH 63
           +Y  LSK+I++ +GG+ N+  V HC TRLRF+L+D  K +  QL+ L  V    ISG Q
Sbjct: 2   DYDKLSKDILQLVGGEENVQRVIHCMTRLRFNLHDNAKADRSQLEQLPGVMGTNISGEQF 61

Query: 64  QIVIGTHVAKVFEEI---NSLIETNSTTKIEQTKKAKAVSRIIDFVSGTFQPILPALSGA 120
           QI+IG  V KV++ I   ++L +  S     Q K    +S + D +SG F PILPA++GA
Sbjct: 62  QIIIGNDVPKVYQAIVRHSNLSDEKSAGSSSQKKNV--LSAVFDVISGVFTPILPAIAGA 119

Query: 121 GMIKALLALLLVFKILTPSSQTYILLNLFADGVFYFLPILIAITAAQKLKANPILALGTV 180
           GMIK L+AL + F  +   SQ +++L    DG FYFLP+L+A++AA+K  +NP +A
Sbjct: 120 GMIKGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLLAMSAARKFGSNPYVAAAIA 179

Query: 181 VMLLHPNWANLVASGKPVSLFHTIPFTLTNYASSVIPIILIICVQAYIEKYLKQIIPKSL 240
             +LHP+   L+ +GKP+S F  +P T   Y+S+VIPI+L I + +Y+EK++ +    SL
Sbjct: 180 AAILHPDLTALLGAGKPIS-FIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASL 238

Query: 241 RLVLVPMLIFLSMGILSFSILGPMGTIAGQYLAVIFTFLSKYASW-APAFLVGAFAPILI 299
           +L++VP    L +  L+   +GP+G I G+YL+    +L  +A   A  FL G F+ ++I
Sbjct: 239 KLIVVPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMIFLAGTFS-LII 297

Query: 300 MFGVHSGIAALGITQLAKLGVDSIFGPGMLCSNIAQATAGTVVTLITKEKKLKEIAGPAA 359
           M G+H     + I  +A+ G D +  P M  +N+ QA A   V L ++ KK K +A   +
Sbjct: 298 MTGMHYAFVPIMINNIAQNGHDYLL-PAMFLANMGQAGASFAVFLRSRNKKFKSLALTTS 356

Query: 360 ITAYMGITEPILYGVNLPKRYPLIASLIGGGLGGLYAGIMNAHRFAV-GSSGLPGLFLYI 418
           ITA MGITEP +YGVN+  + P  A+LIGG  GG + G+     + V G++GLP + ++I
Sbjct: 357 ITALMGITEPAMYGVNMRLKKPFAAALIGGAAGGAFYGMTGVASYIVGGNAGLPSIPVFI 416
```

```
Query: 419 SHTSTHLFITMLIAVIITVSTTAILTF 445
             T  +  I ++IA     S   +L F
Sbjct: 417 GPTFIYAMIGLVIAFAAGTSAAYLLGF 443
```

There is also homology to SEQ ID 2884.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 948

A DNA sequence (GBSx1006) was identified in *S.agalactiae* <SEQ ID 2885> which encodes the amino acid sequence <SEQ ID 2886>. This protein is predicted to be gamma-glutamyl kinase (proB). Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane  160 - 176 ( 160 - 176)

----- Final Results -----
              bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA63147 GB:X92418 gamma-glutamyl kinase [Streptococcus thermophilus]
 Identities = 200/265 (75%), Positives = 235/265 (88%)

Query: 1   MKRHFETTRRIVIKVGTSSLVQTSGKINLSKIDHLAFVISSLMNRGMEVILVSSGAMGFG 60
            MKR+F++ +R+VIK+GTSSLV  SGKINL KID LAFVISSL N+G+EV+LVSSGAMGFG
Sbjct: 1   MKRNFDSVKRLVIKIGTSSLVLPSGKINLEKIDQLAFVISSLHNKGIEVVLVSSGAMGFG 60

Query: 61  LDILKMDKRPQEISQQQAVSSVGQVAMMSLYSQIFSHYQTHVSQILLTRDVVVFPESLQN 120
            L++L ++KRP E+ +QQAVSSVGQVAMMSLYSQ+FSHYQT VSQ+LLTRDVV + ESL N
Sbjct: 61  LNVLDLEKRPAEVGKQQAVSSVGQVAMMSLYSQVFSHYQTKVSQLLLTRDVVEYSESLAN 120

Query: 121 VTNSFESLLSMGILPIVNENDAVSVDEMDHKTKFGDNDRLSAVVAKITKADLLIMLSDID 180
              N+FESL  +G++PIVNENDAVSVDEMDH TKFGDNDRLSA+VAK+  ADLLIMLSDID
Sbjct: 121 AINAFESLFELGVVPIVNENDAVSVDEMDHATKFGDNDRLSAIVAKVVGADLLIMLSDID 180

Query: 181 GLFDKNPNIYDDAVLRSHVSEITDDIIKSAGGAGSKFGTGGMLSKIKSAQMVFDNNGQMI 240
            GLFDKNPN+Y+DA LRS+V EIT++I+ SAGGAGSKFGTGGM+SKIKSAQMVF+N  QM+
Sbjct: 181 GLFDKNPNVYEDATLRSYVPEITEEILASAGGAGSKFGTGGMMSKIKSAQMVFENQSQMV 240

Query: 241 LMNGANPRDILKVLDGHNIGTYFAQ 265
            LMNG NPRDIL+VL+G  IGT F Q
Sbjct: 241 LMNGENPRDILRVLEGAKIGTLFKQ 265
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2887> which encodes the amino acid sequence <SEQ ID 2888>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.97   Transmembrane  163 - 179 ( 163 - 179)
   INTEGRAL    Likelihood = -0.06   Transmembrane  124 - 140 ( 124 - 140)

----- Final Results -----
              bacterial membrane --- Certainty=0.1786(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA63147 GB:X92418 gamma-glutamyl kinase [Streptococcus thermophilus]
 Identities = 212/265 (80%), Positives = 237/265 (89%)

Query: 4   MKRQFEDVTRIVIKIGTSSLVLPTGKINLEKIDQLAFVISSLMNKGKEVILVSSGAMGFG 63
            MKR F+ V R+VIKIGTSSLVLP+GKINLEKIDQLAFVISSL NKG EV+LVSSGAMGFG
Sbjct: 1   MKRNFDSVKRLVIKIGTSSLVLPSGKINLEKIDQLAFVISSLHNKGIEVVLVSSGAMGFG 60

Query: 64  LDILKMEKRPTNLAKQQAVSSVGQVAMMSLYSQIFAYYQTNVSQILLTRDVVVFPESLAN 123
            L++L +EKRP  + KQQAVSSVGQVAMMSLYSQ+F++YQT VSQ+LLTRDVV + ESLAN
Sbjct: 61  LNVLDLEKRPAEVGKQQAVSSVGQVAMMSLYSQVFSHYQTKVSQLLLTRDVVEYSESLAN 120

Query: 124 VTNAFESLISLGIVPIVNENDAVSVDEMDHATKFGDNDRLSAVVAGITKADLLIMLSDID 183
              NAFESL  LG+VPIVNENDAVSVDEMDHATKFGDNDRLSA+VA +  ADLLIMLSDID
Sbjct: 121 AINAFESLFELGVVPIVNENDAVSVDEMDHATKFGDNDRLSAIVAKVVGADLLIMLSDID 180

Query: 184 GLFDKNPTIYEDAQLRSHVANITQEIIASAGGAGSKFGTGGMLSKVQSAQMVFENKGQMV 243
            GLFDKNP +YEDA LRS+V  IT+EI+ASAGGAGSKFGTGGM+SK++SAQMVFEN+ QMV
Sbjct: 181 GLFDKNPNVYEDATLRSYVPEITEEILASAGGAGSKFGTGGMMSKIKSAQMVFENQSQMV 240

Query: 244 LMNGANPRDILRVLEGQPLGTWFKQ 268
            LMNG NPRDILRVLEG  +GT FKQ
Sbjct: 241 LMNGENPRDILRVLEGAKIGTLFKQ 265
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 217/265 (81%), Positives = 242/265 (90%)

Query: 1   MKRHFETTRRIVIKVGTSSLVQTSGKINLSKIDHLAFVISSLMNRGMEVILVSSGAMGFG 60
            MKR FE   RIVIK+GTSSLV  +GKINL KID LAFVISSLMN+G EVILVSSGAMGFG
Sbjct: 4   MKRQFEDVTRIVIKIGTSSLVLPTGKINLEKIDQLAFVISSLMNKGKEVILVSSGAMGFG 63

Query: 61  LDILKMDKRPQEISQQQAVSSVGQVAMMSLYSQIFSHYQTHVSQILLTRDVVVFPESLQN 120
            LDILKM+KRP  +++QQAVSSVGQVAMMSLYSQIF++YQT+VSQILLTRDVVVFPESL N
Sbjct: 64  LDILKMEKRPTNLAKQQAVSSVGQVAMMSLYSQIFAYYQTNVSQILLTRDVVVFPESLAN 123

Query: 121 VTNSFESLLSMGILPIVNENDAVSVDEMDHKTKFGDNDRLSAVVAKITKADLLIMLSDID 180
            VTN+FESL+S+GI+PIVNENDAVSVDEMDH TKFGDNDRLSAVVA ITKADLLIMLSDID
Sbjct: 124 VTNAFESLISLGIVPIVNENDAVSVDEMDHATKFGDNDRLSAVVAGITKADLLIMLSDID 183

Query: 181 GLFDKNPNIYDDAVLRSHVSEITDDIIKSAGGAGSKFGTGGMLSKIKSAQMVFDNNGQMI 240
            GLFDKNP IY+DA LRSHV+ IT +II SAGGAGSKFGTGGMLSK++SAQMVF+N GQM+
Sbjct: 184 GLFDKNPTIYEDAQLRSHVANITQEIIASAGGAGSKFGTGGMLSKVQSAQMVFENKGQMV 243

Query: 241 LMNGANPRDILKVLDGHNIGTYFAQ 265
            LMNGANPRDIL+VL+G  +GT+F Q
Sbjct: 244 LMNGANPRDILRVLEGQPLGTWFKQ 268
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 949

A DNA sequence (GBSx1007) was identified in *S.agalactiae* <SEQ ID 2889> which encodes the amino acid sequence <SEQ ID 2890>. This protein is predicted to be unnamed protein product (proA). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3517(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2891> which encodes the amino acid sequence <SEQ ID 2892>. Analysis of this protein sequence reveals the following:

```
        Possible site: 55
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA63148 GB:X92418 gamma-glutamyl phosphate reductase
          [Streptococcus thermophilus]
 Identities = 309/416 (74%), Positives = 355/416 (85%)

Query: 1   MTDMRRLGQRAKQASLLIAPLSTQIKNRFLSTLAKALVDDTQTLLAANQKDLANAKEHGI 60
           MT +  LGQ+AK AS  IA LST  KN  L+ +AKALV ++  +   N KD+ANA E+GI
Sbjct: 1   MTYVDTLGQQAKVASRQIAKLSTAAKNDLLNQVAKALVAESDYIFTENAKDMANASENGI 60

Query: 61  SDIMMDRLRLTSERIKAIAQGVQQVADLADPIGQVIKGYTNLDGLKILQKRVPLGVIAMI 120
           S IM DRL LT +RI  IA+GV+QVADL DPIGQV++GYTNLDGLKI+QKRVP+GVIAMI
Sbjct: 61  SKIMQDRLLLTEDRIAGIAEGVRQVADLQDPIGQVVRGYTNLDGLKIVQKRVPMGVIAMI 120

Query: 121 FESRPNVSVDAFSLAFKTNNAIILRGGKDALHSNKALVKLIRQSLEKSGITPDAVQLVED 180
           FESRPNVS+DAFSLAFKTNNAIILRGG+DA++SNKALV + R++L+ +GIT DAVQ VED
Sbjct: 121 FESRPNVSIDAFSLAFKTNNAIILRGGRDAINSNKALVTVARKALKNAGITADAVQFVED 180

Query: 181 PSHAVAEELMQATDYVDVLIPRGGAKLIQTVKEKAKVPVIETGVGNVHIYVDAQADLDIA 240
            SH VAEELM AT YVD+LIPRGGA+LIQTVKEKAKVPVIETGVGN HIYVD  A+LD+A
Sbjct: 181 TSHEVAEELMVATKYVDLLIPRGGARLIQTVKEKAKVPVIETGVGNCHIYVDKYANLDMA 240

Query: 241 TKIVINAKTKRPSVCNAAEGLVIHEAVAARFIPMLEKAINQVQPVEWRADDKALPLFEQA 300
           T+IVINAKT+RPSVCNAAE LV+H  +   F+P LEKAI+++Q VE+RAD++AL L E+A
Sbjct: 241 TQIVINAKTQRPSVCNAAESLVVHADIVEEFLPNLEKAISKIQSVEFRADERALKLMEKA 300

Query: 301 VPAKAEDFETEFLDYIMSVKVVSSLEEAISWINQYTSHHSEAIITRDIKAAETFQDLVDA 360
           VPA  EDF TEFLDYIMSVKVV SL+EAI+WIN YT+ HSEAI+T+DI  AE FQD VDA
Sbjct: 301 VPASPEDFATEFLDYIMSVKVVDSLDEAINWINTYTTSHSEAIVTQDISRAEQFQDDVDA 360

Query: 361 AAVYVNASTRFTDGFVFGLGAEIGISTQKMHARGPMGLEALTSTKFYINGDGHIRE 416
           AAVYVNASTRFTDGFVFGLGAEIGISTQKMHARGPMGLEALTSTKFYING G IRE
Sbjct: 361 AAVYVNASTRFTDGFVFGLGAEIGISTQKMHARGPMGLEALTSTKFYINGQGQIRE 416
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 307/417 (73%), Positives = 353/417 (84%), Gaps = 1/417 (0%)

Query: 1   MTYIEILGQNAKKASQSVARLSTASKNEILRDLARNIVADTETILTENARDVVKAKDNGI 60
           MT +  LGQ AK+AS  +A LST  KN  L  LA+ +V DT+T+L  N +D+  AK++GI
Sbjct: 1   MTDMRRLGQRAKQASLLIAPLSTQIKNRFLSTLAKALVDDTQTLLAANQKDLANAKEHGI 60

Query: 61  SEIMVDRLRLNKDRIQAIANGIYQVADLADPIGQVVSGYTNLDGLKILKKRVPLGVIAMI 120
           S+IM+DRLRL  +RI+AIA G+ QVADLADPIGQV+ GYTNLDGLKIL+KRVPLGVIAMI
Sbjct: 61  SDIMMDRLRLTSERIKAIAQGVQQVADLADPIGQVIKGYTNLDGLKILQKRVPLGVIAMI 120

Query: 121 FESRPNVSVDAFSLAFKTGNAIILRGGKDAIFSNTALVNCMRQTLQDTGHNPDIVQLVED 180
           FESRPNVSVDAFSLAFKT NAIILRGGKDA+ SN ALV  +RQ+L+ +G  PD VQLVED
Sbjct: 121 FESRPNVSVDAFSLAFKTNNAIILRGGKDALHSNKALVKLIRQSLEKSGITPDAVQLVED 180

Query: 181 TSHVVAEELMQATDYVDVLIPRGGAKLIQTVKEKSKIPVIETGVGNVHIYIDEFADLDMA 240
            SH VAEELMQATDYVDVLIPRGGAKLIQTVKEK+K+PVIETGVGNVHIY+D  ADLD+A
Sbjct: 181 PSHAVAEELMQATDYVDVLIPRGGAKLIQTVKEKAKVPVIETGVGNVHIYVDAQADLDIA 240

Query: 241 AKIVINAKTQRPSVCNAAEGLVVHQAIAKGFLSQLEKMLKESNQSVEFRADEEALQLLEN 300
            KIVINAKT+RPSVCNAAEGLV+H+A+A  F+  LEK + +  Q VE+RAD++AL L E
```

```
Sbjct: 241 TKIVINAKTKRPSVCNAAEGLVIHEAVAARFIPMLEKAINQV-QPVEWRADDKALPLFEQ 299

Query: 301 AVAASESDYATEFLDYIMSVKVVDSFEQAISWINKYSSHHSEAIITNNISRAEIFQDMVD 360
           AV A   D+ TEFLDYIMSVKVV S E+AISWIN+Y+SHHSEAIIT +I  AE FQD+VD
Sbjct: 300 AVPAKAEDFETEFLDYIMSVKVVSSLEEAISWINQYTSHHSEAIITRDIKAAETFQDLVD 359

Query: 361 AAAVYVNASTRFTDGFVFGLGAEIGISTQKLHARGPMGLEALTSTKYYINGTGQVRE 417
           AAAVYVNASTRFTDGFVFGLGAEIGISTQK+HARGPMGLEALTSTK+YING G +RE
Sbjct: 360 AAAVYVNASTRFTDGFVFGLGAEIGISTQKMHARGPMGLEALTSTKFYINGDGHIRE 416
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 950

A DNA sequence (GBSx1008) was identified in *S.agalactiae* <SEQ ID 2893> which encodes the amino acid sequence <SEQ ID 2894>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1859(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9531> which encodes amino acid sequence <SEQ ID 9532> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2895> which encodes the amino acid sequence <SEQ ID 2896>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0853(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 259/315 (82%), Positives = 287/315 (90%)

Query: 1   MTNDFHHITVLLHETVDMLDIKPDGIYVDATLGGAGHSEYLLSQLGPDGHLYAFDQDQKA 60
           MT +FHH+TVLLHETVDMLDIKPDGIYVDATLGG+GHS YLLS+LG +GHLY FDQDQKA
Sbjct: 22  MTKEFHHVTVLLHETVDMLDIKPDGIYVDATLGGSGHSAYLLSKLGEEGHLYCFDQDQKA 81

Query: 61  IDNAHIRLKKYVDTGQVTFIKDNFRNLSSNLKALGVSEINGICYDLGVSSPQLDERERGF 120
           IDNA + LK Y+D GQVTFIKDNFR+L + L ALGV EI+GI YDLGVSSPQLDERERGF
Sbjct: 82  IDNAQVTLKSYIDKGQVTFIKDNFRHLKARLTALGVDEIDGILYDLGVSSPQLDERERGF 141

Query: 121 SYKQDAPLDMRMNREQSLTAYDVVNTYSYHDLVRIFFKYGEDKFSKQIARKIEQVRAEKT 180
           SYKQDAPLDMRM+R+  LTAY+VVNTY ++DLV+IFFKYGEDKFSKQIARKIEQ RA K
Sbjct: 142 SYKQDAPLDMRMDRQSLLTAYEVVNTYPFNDLVKIFFKYGEDKFSKQIARKIEQARAIKP 201

Query: 181 ISTTTELAEIIKSSKSAKELKKKGHPAKQIFQAIRIEVNDELGAADESIQQAMDLLAVDG 240
           I TTTELAE+IK++K AKELKKKGHPAKQIFQAIRIEVNDELGAADESIQ AM+LLA+DG
Sbjct: 202 IETTTELAELIKAAKPAKELKKKGHPAKQIFQAIRIEVNDELGAADESIQDAMELLALDG 261

Query: 241 RISVITFHSLEDRLTKQLFKEASTVEVPKGLPFIPDDLQPKMELVNRKPILPSQEELEAN 300
           RISVITFHSLEDRLTKQLFKEASTV+VPKGLP IP+D++PK ELV+RKPILPS  EL AN
Sbjct: 262 RISVITFHSLEDRLTKQLFKEASTVDVPKGLPLIPEDMKPKFELVSRKPILPSHSELTAN 321
```

```
Query: 301 NRAHSAKLRVARRIR 315
            RAHSAKLRVA++IR
Sbjct: 322 KRAHSAKLRVAKKIR 336
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 951

A DNA sequence (GBSx1009) was identified in *S.agalactiae* <SEQ ID 2897> which encodes the amino acid sequence <SEQ ID 2898>. This protein is predicted to be FtsL. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.92    Transmembrane   30 -  46 (  24 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC95455 GB:AF068903 YllD [Streptococcus pneumoniae]
 Identities = 44/99 (44%), Positives = 71/99 (71%)

Query: 5    KRTEAVTQTLQRHIKTFSRIEKAFYGAIVITAIIMAVGIIYLQSNSLQVKQEVNQLNSKI 64
            ++ E   Q LQ  +K FSR+EKAFY +I +T +I+A+ II++Q+  LQV+ ++ ++N++I
Sbjct: 3    EKMEKTGQILQMQLKRFSRVEKAFYFSIAVTTLIVAISIIFMQTKLLQVQNDLTKINAQI 62

Query: 65   NDKQTEFDNAKQEVNELSNRDRITKIAKDAGLTIQNDNI 103
             +K+TE D+AKQEVNEL   +R+ +IA    L + N+NI
Sbjct: 63   EEKKTELDDAKQEVNELLRAERLKEIANSHDLQLNNENI 101
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2899> which encodes the amino acid sequence <SEQ ID 2900>. Analysis of this protein sequence reveals the following:

```
     Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.79    Transmembrane   40 -  56 (  37 -  58)

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC95455 GB:AF068903 YllD [Streptococcus pneumoniae]
 Identities = 45/94 (47%), Positives = 69/94 (72%)

Query: 24   LQKRIKTFSRIEKAFYTAIIVTAITMAVSIIYLQSRKLQLQQEITSLNSHISDQKLELNN 83
            LQ ++K FSR+EKAFY +I VT + +A+SII++Q++ LQ+Q ++T +N+ I ++K EL++
Sbjct: 12   LQMQLKRFSRVEKAFYFSIAVTTLIVAISIIFMQTKLLQVQNDLTKINAQIEEKKTELDD 71

Query: 84   AKQEVNELSRRDRIIDIAGKAGLSNRNNNIKKVE 117
            AKQEVNEL R +R+ +IA    L   N NI+   E
Sbjct: 72   AKQEVNELLRAERLKEIANSHDLQLNNENIRIAE 105
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 71/108 (65%), Positives = 87/108 (79%), Gaps = 1/108 (0%)
```

```
Query: 1   MTNEKRTEAVTQTLQRHIKTFSRIEKAFYGAIVITAIIMAVGIIYLQSNSLQVKQEVNQL 60
           MTNEKRT+ VT  LQ+ IKTFSRIEKAFY AI++TAI MAV IIYLQS  LQ++QE+  L
Sbjct: 11  MTNEKRTQVVTNALQKRIKTFSRIEKAFYTAIIVTAITMAVSIIYLQSRKLQLQQEITSL 70

Query: 61  NSKINDKQTEFDNAKQEVNELSNRDRITKIAKDAGLTIQNDNIYRKVD 108
           NS I+D++ E +NAKQEVNELS RDRI  IA  AGL+ +N+NI +KV+
Sbjct: 71  NSHISDQKLELNNAKQEVNELSRRDRIIDIAGKAGLSNRNNNI-KKVE 117
```

SEQ ID 2898 (GBS82) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 2; 2 bands).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 952

A DNA sequence (GBSx1010) was identified in *S.agalactiae* <SEQ ID 2901> which encodes the amino acid sequence <SEQ ID 2902>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1435(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 953

A DNA sequence (GBSx1011) was identified in *S.agalactiae* <SEQ ID 2903> which encodes the amino acid sequence <SEQ ID 2904>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.90    Transmembrane   37 -  53 (  30 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2905> which encodes the amino acid sequence <SEQ ID 2906>. Analysis of this protein sequence reveals the following:

```
     Possible site: 42

>>> Seems to have no N-terminal signal sequence

   INTEGRAL    Likelihood =-13.06    Transmembrane   33 -  49 (  24 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 480/753 (63%), Positives = 603/753 (79%), Gaps = 8/753 (1%)

Query: 5   KKLKKIFLDYVIHIRDRRSPQKNRERVGQNLMILTIFLFFIFIINFVIIVGTDSKFGVNL 64
           KK +K  LDYV+  RDRR+P +NR RVGQN+M+LTIF+FFIFIINF+II+GTD KFGV+L
Sbjct: 2   KKWQKYVLDYVV--RDRRTPVENRVRVGQNMMLLTIFIFFIFIINFMIIIGTDQKFGVSL 59

Query: 65  SKEAKKVYQQSMTVQAKRGTIYDRNGNPIAEDATTYSLYAIISKNYTTATGQKLYVQPSQ 124
           S+ AKKVYQ+++T+QAKRGTIYDRNG  IA D+TTYS+YAI+ K++ +A+ +KLYVQPSQ
Sbjct: 60  SEGAKKVYQETVTIQAKRGTIYDRNGTAIAVDSTTYSIYAILDKSFVSASDEKLYVQPSQ 119

Query: 125 YEKVASILENKLGMKKNLVLKQLNQKKLFQVSFGSSGSGLSYTKMADIKKTMEKSDIKGI 184
           YE VA IL+  LGMKK  V+KQL +K LFQVSFG SGSG+SY+ M+ I+K ME + IKGI
Sbjct: 120 YETVADILKKHLGMKKTDVIKQLKRKGLFQVSFGPSGSGISYSTMSTIQKAMEDAKIKGI 179

Query: 185 GFSTSPGRIYPNGIFASQFIGF-TLPQDDGDG-KKLVGNTGLEAALNKVLSGTDGKVTYE 242
            F+TSPGR+YPNG FAS+FIG  +L +D   G K LVG TGLEA+ +K+LSG DG +TY+
Sbjct: 180 AFTTSPGRMYPNGTFASEFIGLASLTEDKKTGVKSLVGKTGLEASFDKILSGQDGVITYQ 239

Query: 243 KDRSGNVLLGTATTERRAVNGKDIYTTLSEPIQTVLETQMDVFAEKTKGKFASATVVNAK 302
           KDR+G  LLGT  T ++A++GKDIYTTLSEPIQT LETQMDVF  K+ G+ ASAT+VNAK
Sbjct: 240 KDRNGTTLLGTGKTVKKAIDGKDIYTTLSEPIQTFLETQMDVFQAKSNGQLASATLVNAK 299

Query: 303 TGEILATSQRPTYNPSTLKGYDKKNLGTYNTLLYDNFFEPGSTMKVMTLASAIDSKHFNS 362
           TGEILAT+QRPTYN  TLKG +  N   Y+ L   N FEPGSTMKVMTLA+AID K FN
Sbjct: 300 TGEILATTQRPTYNADTLKGLENTNYKWYSALHQGN-FEPGSTMKVMTLAAAIDDKVFNP 358

Query: 363 TEVYNSAQ-YKIADAIIRDWDVNEGLSSGSYMTFPQGFAHSSNVGMVTLEQKMGRDKWLN 421
            E +++A    IADA I+DW +NEG+S+G YM + QGFA SSNVGM  LEQKMG  KW+N
Sbjct: 359 NETFSNANGLTIADATIQDWSINEGISTGQYMNYAQGFAFSSNVGMTKLEQKMGNAKWMN 418

Query: 422 YLSKFKFGYPTRFGMLHESGGLFPSDNEVTIAMSSFGQGIGVTQVQMLRAFTSISNDGVM 481
           YL+KF+FG+PTRFG+  E  G+FPSDN VT AMS+FGQGI VTQ+QMLRAFT+ISN+G M
Sbjct: 419 YLTKFRFGFPTRFGLKDEDAGIFPSDNIVTQAMSAFGQGISVTQIQMLRAFTAISNNGEM 478

Query: 482 LQPQFISSIYDPNTGTSRTARKEVVGKPVSKEAASKTRDYMVTVGTDPYYGTLYA-AGAP 540
           L+PQFIS IYDPNT + RTA KE+VGKPVSK+AAS+TR YM+ VGTDP +GTLY+    P
Sbjct: 479 LEPQFISQIYDPNTASFRTANKEIVGKPVSKKAASETRQYMIGVGTDPEFGTLYSKTFGP 538

Query: 541 VIQVGNQSVAVKSGTAQIAQEGGGGYLQ-GKNDTINSVVAMVPSENPDFIMYVTIQQPEK 599
           +I+VG+  VAVKSGTAQI  E G GY   G  + + SVVAMVP++ PDF+MYVT+ +P+
Sbjct: 539 IIKVGDLPVAVKSGTAQIGSEDGSGYQDGGLTNYVYSVVAMVPADKPDFLMYVTMTKPQH 598

Query: 600 FSITFWKDVVNPVLEQATAMKETILKPGLNDSEHQTKYKLSKIVGENPGHVAEELRRNLV 659
           F   FW+DVVNPVLE+A  M++T+ KP ++D+  QT YKL   VG+NPG  + ELRRNLV
Sbjct: 599 FGPLFWQDVVNPVLEEAYLMQDTLTKPVVSDANRQTTYKLPNFVGKNPGETSSELRRNLV 658

Query: 660 QPIILGNGSKVSKVSKRPGANLAENEQLLVLTNKLTELPDMYGWSKANVEQFAKWTGIKV 719
           QP++LG GSK+ KVS +PG  L EN+Q+L+L+++  E+PDMYGW+K+NV+ FAKWTGI +
Sbjct: 659 QPVVLGTGSKIKKVSHQPGQTLTENQQVLILSDRFVEVPDMYGWTKSNVKTFAKWTGIDI 718

Query: 720 TYKGSTSGKVRKQSIDVGKSINKIKKIKITIGD 752
           ++KG+ SG+V KQS+DVGKS+ KIKK+ IT+GD
Sbjct: 719 SFKGTDSGRVMKQSVDVGKSLKKIKKMTITLGD 751
```

A related GBS gene <SEQ ID 8691> and protein <SEQ ID 8692> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -4.31
GvH: Signal Score (-7.5): -7.07
     Possible site: 47
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -13.90 threshold:  0.0
```

```
  INTEGRAL    Likelihood =-13.90   Transmembrane   37 -  53 (  30 -  60)
  PERIPHERAL  Likelihood =  5.30     450
 modified ALOM score:   3.28

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00411(301 - 2556 of 2856)
GP|6779111|emb|CAB70457.1||A94911(1 - 752 of 752) unnamed protein product {unidentified},
homology to penicillin-binding protein 2x (S. pneumoniae)
%Match = 77.4
%Identity = 99.7  %Similarity = 99.9
Matches = 750  Mismatches = 1  Conservative Sub.s = 1

66        96        126       156       186       216       246       276
RIEKAFYGAIVITAIIMAVGIIYLQSNSLQVKQEVNQLNSKINDKQTEFDNAKQEVNELSNRDRITKIAKDAGLTIQNDN

306       336       366       396       426       456       486       516
IYRKVD*SVTFFKKLKKIFLDYVIHIRDRRSPQKNRERVGQNLMILTIFLFFIFIINFVIIVGTDSKFGVNLSKEAKKVY
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        VTFFKKLKKIFLDYVIHIRDRRSPQKNRERVGQNLMILTIFLFFIFIINFVIIVGTDSKFGVNLSKEAKKVY
                10        20        30        40        50        60        70

546       576       606       636       666       696       726       756
QQSMTVQAKRGTIYDRNGNPIAEDATTYSLYAIISKNYTTATGQKLYVQPSQYEKVASILENKLGMKKNLVLKQLNQKKL
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QQSMTVQAKRGTIYDRNGNPIAEDATTYSLYAIISKNYTTATGQKLYVQPSQYEKVASILENKLGMKKNLVLKQLNQKKL
                90       100       110       120       130       140       150

786       816       846       876       906       936       966       996
FQVSFGSSGSGLSYTKMADIKKTMEKSDIKGIGFSTSPGRIYPNGIFASQFIGFTLPQDDGDGKKLVGNTGLEAALNKVL
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
FQVSFGSSGSGLSYTKMADIKKTMEKSDIKGIGFSTSPGRIYPNGIFASQFIGFTLPQDDGDGKKLVGNTGLEAALNKVL
               170       180       190       200       210       220       230

1026      1056      1086      1116      1146      1176      1206      1236
SGTDGKVTYEKDRSGNVLLGTATTERRAVNGKDIYTTLSEPIQTVLETQMDVFAEKTKGKFASATVVNAKTGEILATSQR
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGTDGKVTYEKDRSGNVLLGTATTERRAVNGKDIYTTLSEPIQTVLETQMDVFAEKTKGKFASATVVNAKTGEILATSQR
               250       260       270       280       290       300       310

1266      1296      1326      1356      1386      1416      1446      1476
PTYNPSTLKGYDKKNLGTYNTLLYDNFFEPGSTMKVMTLASAIDSKHFNSTEVYNSAQYKIADAIIRDWDVNEGLSSGSY
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
PTYNPSTLKGYDKKNLGTYNTLLYDNFFEPGSTMKVMTLASAIDSKHFNSTEVYNSAQYKIADAVIRDWDVNEGLSSGSY
               330       340       350       360       370       380       390

1506      1536      1566      1596      1626      1656      1686      1716
MTFPQGFAHSSNVGMVTLEQKMGRDKWLNYLSKFKFGYPTRFGMLHESGGLFPSDNEVTIAMSSFGQGIGVTQVQMLRAF
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MTFPQGFAHSSNVGMVTLEQKMGRDKWLNYLSKFKFGYPTRFGMLHESGGLFPSDNEVTIAMSSFGQGIGVTQVQMLRAF
               410       420       430       440       450       460       470

1746      1776      1806      1836      1866      1896      1926      1956
TSISNDGVMLQPQFISSIYDPNTGTSRTARKEVVGKPVSKEAASKTRDYMVTVGTDPYYGTLYAAGAPVIQVGNQSVAVK
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TSISNDGVMLQPQFISSIYDPNTGTSRTARKEVVGKPVSKEAASKTRDYMVTVGTDPYYGTLYAAGAPVIQVGNQSVAVK
               490       500       510       520       530       540       550

1986      2016      2046      2076      2106      2136      2166      2196
SGTAQIAQEGGGGYLQGKNDTINSVVAMVPSENPDFIMYVTIQQPEKFSITFWKDVVNPVLEQATAMKETILKPGLNDSE
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
```

```
SGTAQIAQEGGGGYLQGKNDTINSVVAMVPSENPDFIMYVTIQQPEKFSITFWKDVVNPVLEQATAMKETILKPVLNDSE
               570       580       590       600       610       620       630

2226      2256      2286      2316      2346      2376      2406      2436
HQTKYKLSKIVGENPGHVAEELRRNLVQPIILGNGSKVSKVSKRPGANLAENEQLLVLTNKLTELPDMYGWSKANVEQFA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HQTKYKLSKIVGENPGHVAEELRRNLVQPIILGNGSKVSKVSKRPGANLAENEQLLVLTNKLTELPDMYGWSKANVEQFA
               650       660       670       680       690       700       710

2466      2496      2526      2556      2586      2616      2646      2676
KWTGIKVTYKGSTSGKVRKQSIDVGKSINKIKKIKITIGD*HVFKYNGRCHSICPDSYCHSALH*VLPIEENWRATNA*R
||||||||||||||||||||||||||||||||||||||||
KWTGIKVTYKGSTSGKVRKQSIDVGKSINKIKKIKITIGD
               730       740       750
```

SEQ ID 8692 (GBS352d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 15 & 16; MW 105.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 17 & 18; MW 80.5kDa), in Figure 182 (lane 3; MW 80kDa) and in Figure 185 (lane 4; MW 105kDa). Purified GBS352d-GST is shown in lane 5 of Figure 236.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 954

A DNA sequence (GBSx1012) was identified in *S.agalactiae* <SEQ ID 2907> which encodes the amino acid sequence <SEQ ID 2908>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1950(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 955

A DNA sequence (GBSx1013) was identified in *S.agalactiae* <SEQ ID 2909> which encodes the amino acid sequence <SEQ ID 2910>. This protein is predicted to be unnamed protein product (mraY). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-15.12    Transmembrane   56 -  72 (  47 -  76)
   INTEGRAL    Likelihood =-14.70    Transmembrane  203 - 219 ( 198 - 223)
   INTEGRAL    Likelihood = -6.69    Transmembrane  318 - 334 ( 315 - 335)
   INTEGRAL    Likelihood = -6.64    Transmembrane   83 -  99 (  79 - 103)
   INTEGRAL    Likelihood = -5.52    Transmembrane  179 - 195 ( 175 - 197)
   INTEGRAL    Likelihood = -5.31    Transmembrane  232 - 248 ( 230 - 249)
   INTEGRAL    Likelihood = -3.08    Transmembrane  119 - 135 ( 119 - 137)
   INTEGRAL    Likelihood = -2.87    Transmembrane  151 - 167 ( 147 - 167)
   INTEGRAL    Likelihood = -2.34    Transmembrane  254 - 270 ( 254 - 270)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2911> which encodes the amino acid sequence <SEQ ID 2912>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.55   Transmembrane   52 -  68 (  48 -  75)
   INTEGRAL    Likelihood = -9.39   Transmembrane  175 - 191 ( 171 - 194)
   INTEGRAL    Likelihood = -8.12   Transmembrane   30 -  46 (  23 -  48)
   INTEGRAL    Likelihood = -6.37   Transmembrane  121 - 137 ( 119 - 145)
   INTEGRAL    Likelihood = -6.32   Transmembrane  293 - 309 ( 287 - 309)
   INTEGRAL    Likelihood = -5.31   Transmembrane  204 - 220 ( 202 - 221)
   INTEGRAL    Likelihood = -5.20   Transmembrane  151 - 167 ( 150 - 170)
   INTEGRAL    Likelihood = -4.67   Transmembrane  226 - 242 ( 224 - 244)
   INTEGRAL    Likelihood = -0.11   Transmembrane   91 - 107 (  91 - 107)

----- Final Results -----
              bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB70458 GB:A94911 unnamed protein product [unidentified]
 Identities = 244/309 (78%), Positives = 273/309 (87%), Gaps = 1/309 (0%)

Query: 1   LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLLVATAVSLLVSLF-SIKNTQSLALISGIL 59
            LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFL+VA  VSL+ S+  S +N+ +L    GIL
Sbjct: 28  LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLVVALLVSLIFSIILSKENSGNLGATFGIL 87

Query: 60  SIVVIYGIIGFLDDFLKIFKQINEGLTAKQKLALQLVGGLMFYFLHVSPSGISSINVFGY 119
            S+V+IYGIIGFLDDFLKIFKQINEGLT KQK++LQL+ GL+FYF+HV PSG S+IN+FG+
Sbjct: 88  SVVLIYGIIGFLDDFLKIFKQINEGLTPKQKMSLQLIAGLIFYFVHVLPSGTSAINIFGF 147

Query: 120 QLPLGIFYLFFVLFWVVGFSNAVNLTDGIDGLASISVVISLVTYGVIAYVQSQFDVLLLI 179
             L +G  Y FFVLFWVVGFSNAVNLTDGIDGLASISVVISL+TYG+IAY Q+QFD+LL+I
Sbjct: 148 NLEVGYLYAFFVLFWVVGFSNAVNLTDGIDGLASISVVISLITYGIIAYNQTQFDILLII 207

Query: 180 GAMIGALLGFFCFNHKPAKVFMGDVGSLALGAMLAAISIALRQEWTLLIIGIVYVLETSS 239
              MIGALLGFF FNHKPAKVFMGDVGSLALGAMLAAISIALRQEWTLL IG VYV ETSS
Sbjct: 208 VIMIGALLGFFVFNHKPAKVFMGDVGSLALGAMLAAISIALRQEWTLLFIGFVYVFETSS 267

Query: 240 VMLQVSYFKYTKKKYGEGRRIFRMTPFHHHLELGGLSGKGKKWSEWQVDAFLWGVGSLAS 299
            VMLQV+YFKYTKKK G G+RIFRMTPFHHHLELGG+SGKG KWSEW+VDAFLW +G   S
Sbjct: 268 VMLQVAYFKYTKKKTGVGKRIFRMTPFHHHLELGGVSGKGNKWSEWKVDAFLWAIGIFMS 327

Query: 300 LLVLAILYV 308
             + LAILY+
Sbjct: 328 AITLAILYL 336
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 244/309 (78%), Positives = 273/309 (87%), Gaps = 1/309 (0%)

Query: 28  LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLIVALLVSLIFSIILSKENSGNLGATFGIL 87
            LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFL+VA  VSL+ S+  S +N+ +L    GIL
Sbjct: 1   LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLLVATAVSLLVSLF-SIKNTQSLALISGIL 59

Query: 88  SVVLIYGIIGFLDDFLKIFKQINEGLTPKQKMSLQLIAGLIFYFVHVLPSGTSAINIFGF 147
            S+V+IYGIIGFLDDFLKIFKQINEGLT KQK++LQL+ GL+FYF+HV PSG S+IN+FG+
Sbjct: 60  SIVVIYGIIGFLDDFLKIFKQINEGLTAKQKLALQLVGGLMFYFLHVSPSGISSINVFGY 119
```

```
Query: 148 YLEVGYLYAFFVLFWVVGFSNAVNLTDGIDGLASISVVISLITYGIIAYNQTQFDILLII 207
            L +G  Y FFVLFWVVGFSNAVNLTDGIDGLASISVVISL+TYG+IAY Q+QFD+LL+I
Sbjct: 120 QLPLGIFYLFFVLFWVVGFSNAVNLTDGIDGLASISVVISLVTYGVIAYVQSQFDVLLLI 179

Query: 208 VIMIGALLGFFVFNHKPAKVFMGDVGSLALGAMLAAISIALRQEWTLLFIGFVYVFETSS 267
             MIGALLGFF FNHKPAKVFMGDVGSLALGAMLAAISIALRQEWTLL IG VYV ETSS
Sbjct: 180 GAMIGALLGFFCFNHKPAKVFMGDVGSLALGAMLAAISIALRQEWTLLIIGIVYVLETSS 239

Query: 268 VMLQVAYFKYTKKKTGVGKRIFRMTPFHHHLELGGVSGKGNKWSEWKVDAFLWAIGIFMS 327
           VMLQV+YFKYTKKK G G+RIFRMTPFHHHLELGG+SGKG KWSEW+VDAFLW +G   S
Sbjct: 240 VMLQVSYFKYTKKKYGEGRRIFRMTPFHHHLELGGLSGKGKKWSEWQVDAFLWGVGSLAS 299

Query: 328 AITLAILYL 336
            + LAILY+
Sbjct: 300 LLVLAILYV 308
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 956

A DNA sequence (GBSx1014) was identified in *S.agalactiae* <SEQ ID 2913> which encodes the amino acid sequence <SEQ ID 2914>. This protein is predicted to be autoaggregation-mediating protein (deaD). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3018(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14444 GB:Z99116 similar to ATP-dependent RNA helicase
           [Bacillus subtilis]
 Identities = 215/436 (49%), Positives = 310/436 (70%), Gaps = 5/436 (1%)

Query: 3   FKDFNFKPYIQRALDELKFVDPTDVQAKLIPVVRSGRDLVGESKTGSGKTHTFLLPIFEK 62
           F+ +  KP+I  A+  L F +PTD+Q +LIP V     ++G+S+TG+GKTH +LLP+  K
Sbjct: 6   FELYELKPFIIDAVHRLGFYEPTDIQKRLIPAVLKKESVIGQSQTGTGKTHAYLLPLLNK 65

Query: 63  LDESSDDVQVVITAPSRELGTQIYQATKQIAEHSE-QEIRVVNYVGGTDKLRQIEKLKVS 121
           +D + D VQVVITAP+REL  QIYQ   +I +  E  +IR   ++GGTDK + I+KLK+
Sbjct: 66  IDPAKDVVQVVITAPTRELANQIYQEALKITQGEEGSQIRSKCFIGGTDKQKSIDKLKI- 124

Query: 122 QPHIVIGTPGRIYDLVKSGDLAIHKAHTFVVDEADMTLDMGFLDTVDKIAGSLPKDVQIL 181
           QPH+V+GTPGRI DL+K   L++HKA + V+DEAD+ LDMGFL  VD I   +P+D+Q+L
Sbjct: 125 QPHLVVGTPGRIADLIKEQALSVHKAESLVIDEADLMLDMGFLADVDYIGSRMPEDLQML 184

Query: 182 VFSATIPQKLQPFLKKYLTNPVMEKIKTATVIADTIDNWLLSTKGRDKNAQILELSKLMQ 241
           VFSATIP+KL+PFLKKY+ NP    ++   V A  I++ L+ +K RDK+  + ++   +
Sbjct: 185 VFSATIPEKLKPFLKKYMENPKYAHVEPKQVTAAKIEHILIPSKHRDKDKLLFDIMSHLN 244

Query: 242 PYLAMIFVNTKERADELHSYLSSNGLKVAKIHGGIAPRERKRIMNQVKNLEFEYIVATDL 301
           PYL ++F NTK  AD +  YL+  G+K+  +HGG+ PRERK++M Q+ +LEF YI+ATDL
Sbjct: 245 PYLGIVFANTKNTADHIAQYLTGKGMKIGLLHGGLTPRERKKVMKQINDLEFTYIIATDL 304

Query: 302 AARGIDIEGVSHVINDAIPQDLSFFVHRVGRTGRNGLSGTAITLYQPSDDSDIRELEKLG 361
           AARGIDI+GVSHVIN  +P DL F+VHRVGRT R G SG A+T+Y+ +D+  +  LEK+G
Sbjct: 305 AARGIDIKGVSHVINYELPDDLDFYVHRVGRTARAGSSGQAMTIYELTDEDALVRLEKMG 364

Query: 362 INFIPKVIKNGEFQDTYDRDRRNNREKSYQKLDTEMIGLVKKKKKKIKPGYKKKIQWKVD 421
           I F    ++ GE++   DR RR  R+K+  + D E+   + KK KK+KPGYKKK+ ++++
```

```
Sbjct: 365 IEFEYLELEKGEWKKGDDRQRRKKRKKTPNEAD-EIAHRLVKKPKKVKPGYKKKMSYEME 423

Query: 422 EKRRKERRASNRAKGR 437
           + ++K+RR  N++K R
Sbjct: 424 KIKKKQRR--NQSKKR 437
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2915> which encodes the amino acid sequence <SEQ ID 2916>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2315(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 382/447 (85%), Positives = 420/447 (93%)

Query: 1   MSFKDFNFKPYIQRALDELKFVDPTDVQAKLIPVVRSGRDLVGESKTGSGKTHTFLLPIF 60
            MSFKD++FK Y+Q+AL+E+ FV+PT+VQ +LIP+V SGRDLVGESKTGSGKTHTFLLPIF
Sbjct: 1   MSFKDYHFKQYVQQALEEIGFVNPTEVQKRLIPIVNSGRDLVGESKTGSGKTHTFLLPIF 60

Query: 61  EKLDESSDDVQVVITAPSRELGTQIYQATKQIAEHSEQEIRVVNYVGGTDKLRQIEKLKV 120
            EKLDE+  +VQVVITAPSREL TQI+ A KQIA+H ++EIR+ NYVGGTDKLRQIEKLK
Sbjct: 61  EKLDEAKAEVQVVITAPSRELATQIFDACKQIAKHFQEEIRLANYVGGTDKLRQIEKLKD 120

Query: 121 SQPHIVIGTPGRIYDLVKSGDLAIHKAHTFVVDEADMTLDMGFLDTVDKIAGSLPKDVQI 180
            SQPHIVIGTPGRIYDLVKSGDLAIHKA TFVVDEADMT+DMGFLDTVDKIA SLPK VQI
Sbjct: 121 SQPHIVIGTPGRIYDLVKSGDLAIHKATTFVVDEADMTMDMGFLDTVDKIAASLPKSVQI 180

Query: 181 LVFSATIPQKLQPFLKKYLTNPVMEKIKTATVIADTIDNWLLSTKGRDKNAQILELSKLM 240
            LVFSATIPQKLQPFLKKYLTNPV+E+IKT TVIADTIDNWL+STKGRDKN Q+LE+ K M
Sbjct: 181 LVFSATIPQKLQPFLKKYLTNPVIEQIKTKTVIADTIDNWLVSTKGRDKNGQLLEILKTM 240

Query: 241 QPYLAMIFVNTKERADELHSYLSSNGLKVAKIHGGIAPRERKRIMNQVKNLEFEYIVATD 300
            QPY+AM+FVNTKERAD+LH++L++NGLKVAKIHGGI PRERKRIMNQVK L+FEYIVATD
Sbjct: 241 QPYMAMLFVNTKERADDLHAFLTANGLKVAKIHGGIPPRERKRIMNQVKKLDFEYIVATD 300

Query: 301 LAARGIDIEGVSHVINDAIPQDLSFFVHRVGRTGRNGLSGTAITLYQPSDDSDIRELEKL 360
            LAARGIDIEGVSHVINDAIPQDLSFFVHRVGRTGRNG++GTAITLYQPSDDSDI+ELEK+
Sbjct: 301 LAARGIDIEGVSHVINDAIPQDLSFFVHRVGRTGRNGMAGTAITLYQPSDDSDIKELEKM 360

Query: 361 GINFIPKVIKNGEFQDTYDRDRRNNREKSYQKLDTEMIGLVKKKKKKIKPGYKKKIQWKV 420
            GI F PKV+KNGEFQDTYDRDRR NREK+YQKLDTEMIGLVKKKKKK+KPGYKKKIQW V
Sbjct: 361 GIAFTPKVLKNGEFQDTYDRDRRQNREKAYQKLDTEMIGLVKKKKKKVKPGYKKKIQWAV 420

Query: 421 DEKRRKERRASNRAKGRAERKAKKQSF 447
            DEKRRKERRA NRAKGRAERKAKKQ F
Sbjct: 421 DEKRRKERRAENRAKGRAERKAKKQHF 447
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 957

A DNA sequence (GBSx1015) was identified in *S.agalactiae* <SEQ ID 2917> which encodes the amino acid sequence <SEQ ID 2918>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
```

```
----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 2920.

A related GBS gene <SEQ ID 8693> and protein <SEQ ID 8694> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      8.85
GvH: Signal Score (-7.5): -1.77
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 0 value:  8.12 threshold:  0.0
   PERIPHERAL  Likelihood =  8.12     182
 modified ALOM score:  -2.12

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
EGAD|126750| collagen binding protein Insert characterized
  GP|1617328|emb|CAA68052.1||X99716 collagen binding protein Insert characterized

ORF00181(331 - 1089 of 1410)
EGAD|126750|135177(23 - 260 of 263) collagen binding protein {Lactobacillus
reuteri}GP|1617328|emb|CAA68052.1||X99716 collagen bindi
ng protein {Lactobacillus reuteri}
%Match = 11.2
%Identity = 35.4  %Similarity = 59.0
Matches = 69  Mismatches = 77  Conservative Sub.s = 46

177       207       237       267       297       327       357       387
KTKFLKLLKSEISSFQAFLLI*NLYHLIRKYYYTDRF*SVRLVI*YFRRILMFKKIILSIATIAATASLAVSVQASEKVE
                                       :::    :  |  :     :  |  : || ||: :   | |
                                       MKFWKKALLTIAALTVGTSAGITSVSAASSAVNSELVHKGE
                                               10        20        30        40

417       447       477       507       537       567       597       627
LKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTISTGIDAGKFDLSANDFSYNKERAEKYLFSDP
| :  :   :|::|:|: |: |::||: ||| |    |  |    :|::  |: :||||:  |: :   |||::| || |
LTIGLEGTYSPYSYRKNNKLTGFEVDLGKAVAKKMGLKANFVPTKWDSLIAGLGSGKFDVVMNNITQTPERAKQYNFSTP
                 60        70        80        90       100       110       120

657       687       717       747
XSRSNYAVVGKKGSHYKSLSDLSGKSTEVLSGVNYAQVLENWNKN-HPN-------------------------------
  :| :|::    |: ||| |: ||     :| | | |:: :  |  ||
YIKSRFALIVPTDSNIKSLKDIKGKKIIAGTGTNNANVVKKYKGNLTPNGDFASSLDMIKQGRAAGTVNSREAWYAYSKK
                140       150       160       170       180       190       200

789       819       849       879       909       939       969
--------------KKPIKIKYVSGTTGVTSRLKNIESGKIDFILYDAISSDYIVKDQSLNLSVSPLKGKIGNNKDGLEY
              : | ||  :
NSTKGLKMIDVSSEQDPAKISALF--------------------------------------------------------
                220

999      1029      1059      1089      1119      1149      1179      1209
LLLPKDKKGKTLQKFINKRIKVLKENGTLARLSKQYFGGDYVSNIDK*ISETISFIFLHVRVLRDRITEIESLEKESRRN
     :||   :|   || :| |:::||: :||::||| |
-----NKKDTAIQSSYNKALKELQQDGTVKKLSEKYFGADITE
```

```
        230       240       250       260
```

SEQ ID 8694 (GBS8) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 5; MW 31kDa), Figure 63 (lane 2; MW 31.3kDa), Figure 66 (lane 2 & 3; MW 31kDa), in Figure 178 (lane 2; MW 31kDa), in Figure 179 (lane 3 & 4; MW 31kDa) and in Figure 180 (lane 3; MW 31kDa). It was also expressed in *E.coli* as a GST-fusion product, with SDS-PAGE shown in Figure 66 (lanes 4 & 5; MW 56kDa) and in Figure 180 (lanes 4 & 5; MW 55kDa).

GBS8-His was purified as shown in Figures 189 (lane 7), 211 (lane 3), 228 (lanes 4-5) and 230 (lanes 3-6). Purified GBS8-GST is shown in Figure 209, lane 6.

The GBS8-His fusion product was purified (Figure 90A) and used to immunise mice (lane 2 product; 12.9μg/mouse). The resulting antiserum was used for Western blot (Figure 90B), FACS (Figure 90C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 958

A DNA sequence (GBSx1016) was identified in *S.agalactiae* <SEQ ID 2921> which encodes the amino acid sequence <SEQ ID 2922>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3991(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 959

A DNA sequence (GBSx1017) was identified in *S.agalactiae* <SEQ ID 2923> which encodes the amino acid sequence <SEQ ID 2924>. This protein is predicted to be probable amino-acid abc transporter permease protein in idh-deor inter. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.62   Transmembrane   50 -  66 (  41 -  74)
   INTEGRAL    Likelihood = -0.90   Transmembrane  226 - 242 ( 226 - 242)
   INTEGRAL    Likelihood = -0.53   Transmembrane   80 -  96 (  80 -  96)

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15985 GB:Z99124 similar to amino acid ABC transporter
          (permease) [Bacillus subtilis]
 Identities = 90/224 (40%), Positives = 137/224 (60%), Gaps = 10/224 (4%)

Query: 28  WKAVLDAIPSILERLPITLLLTVAGALFGLILALIFAVVKINRVKILYPIQALFVSFLRG 87
            W+ ++ A P++++ LPITL + +A  +F +I  LI A++  N++ +L+ +  L++SF RG
Sbjct: 6   WEFMISAFPTLIQALPITLFMAIAAMIFAIIGGLILALITKNKIPVLHQLSKLYISFFRG 65

Query: 88  TPILVQLMLSYYGIPLFLKFLNQKYGFDWNINAIPASVFAITAFAFNEAAYTSETIRAAI 147
             P LVQL L YYG+P     +++          + A   AI   +   AAY +E  RAA+
Sbjct: 66  VPTLVQLFLIYYGLPQLFPEMSK----------MTALTAAIIGLSLKNAAYLAEIFRAAL 115

Query: 148 LSVDQGEIEAARSLGMTSAQVYRRVIIPNAAVVATPTLINTLIGLTKGTSLAFNAGIVEM 207
             SVD G++EA  S+GMT  Q YRR+I+P A   A P   NT IGL K TSLAF  G++EM
Sbjct: 116 NSVDDGQLEACLSVGMTKFQAYRRIILPQAIRNAIPATGNTFIGLLKETSLAFTLGVMEM 175

Query: 208 FAQAQIMGGSDYRYFERYISVALVYWAVSFLIEQLGNAIERKMA 251
            FAQ ++    + +YFE Y++VA+VYW ++ +   L +  ER M+
Sbjct: 176 FAQGKMYASGNLKYFETYLAVAIVYWVLTIIYSILQDLFERAMS 219
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2925> which encodes the amino acid sequence <SEQ ID 2926>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.27   Transmembrane   80 -  96 (  74 - 104)
   INTEGRAL    Likelihood = -1.06   Transmembrane  207 - 223 ( 207 - 223)
   INTEGRAL    Likelihood = -0.90   Transmembrane  110 - 126 ( 110 - 126)

----- Final Results -----
               bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9167> which encodes the amino acid sequence <SEQ ID 9168>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -7.27   Transmembrane   50 -  66 (  44 -  74)
     INTEGRAL    Likelihood = -1.06   Transmembrane  177 - 193 ( 177 - 193)
     INTEGRAL    Likelihood = -0.90   Transmembrane   80 -  96 (  80 -  96)

----- Final Results -----
               bacterial membrane --- Certainty=0.391(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 212/267 (79%), Positives = 238/267 (88%)

Query: 1    MNQFILTGGWSWYNNLVSQVPAGKLFSWKAVLDAIPSILERLPITLLLTVAGALFGLILA 60
            M    LT GW++Y+ L+S +P GKLFSW AV DAIP+I++RLPITL LT++GA FGL+LA
Sbjct: 31   MTSVFLTSGWAFYDYLISPIPHGKLFSWHAVFDAIPNIIQRLPITLGLTLSGATFGLVLA 90

Query: 61   LIFAVVKINRVKILYPIQALFVSFLRGTPILVQLMLSYYGIPLFLKFLNQKYGFDWNINA 120
            LIFA+VKIN+VK+LYPIQA+FVSFLRGTPILVQLML+YYGIPLFLKFLNQKYGFDWN+NA
Sbjct: 91   LIFALVKINKVKLLYPIQAIFVSFLRGTPILVQLMLTYYGIPLFLKFLNQKYGFDWNVNA 150

Query: 121 IPASVFAITAFAFNEAAYTSETIRAAILSVDQGEIEAARSLGMTSAQVYRRVIIPNAAVV 180
            IPAS+FAITAFAFNEAAY SETIRAAILSVD GEIEAA+SLGMTS QVYRRVIIPNA VV
Sbjct: 151 IPASIFAITAFAFNEAAYASETIRAAILSVDTGEIEAAKSLGMTSVQVYRRVIIPNATVV 210

Query: 181 ATPTLINTLIGLTKGTSLAFNAGIVEMFAQAQIMGGSDYRYFERYISVALVYWAVSFLIE 240
            A PTLIN LIGLTKGTSLAFNAGIVEMFAQAQI+GGSDYRYFERYISVALVYW++S L+E
```

```
Sbjct: 211 AIPTLINGLIGLTKGTSLAFNAGIVEMFAQAQILGGSDYRYFERYISVALVYWSISILME 270

Query: 241 QLGNAIERKMAIKAPRHLTDEIPGGVR 267
           Q+G  IE KMAIKAP    +E  G +R
Sbjct: 271 QVGRLIENKMAIKAPEQARNEKLGELR 297
```

There is also homology to SEQ ID 4794.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 960

A DNA sequence (GBSx1018) was identified in *S.agalactiae* <SEQ ID 2927> which encodes the amino acid sequence <SEQ ID 2928>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3205(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00329 GB:AF008220 putative amino acid transporter [Bacillus subtilis]
 Identities = 121/247 (48%), Positives = 176/247 (70%)

Query: 1   MIKLRQLTKSFSGQKVLDKLDLDIEKGQVVALVGASGAGKSTFLRSMNYLEEPDYGTIEI 60
           MI+++ + K F    VL  ++L + KG+VV ++G SG+GK+TFLR +N LE PD G I I
Sbjct: 1   MIEIKNIHKQFGIHHVLKGINLTVRKGEVVTIIGPSGSGKTTFLRCLNLLERPDEGIISI 60

Query: 61  DDFKVDFKSISKDDILTLRRKLAMVFQQFNLFERRTALDNVKEGLKIVKKMSDQEATRIA 120
            D  ++ +  SK ++  LR++ AMVFQQ++LF  +T ++NV EGL I +KM  Q+A  +A
Sbjct: 61  HDKVINCRFPSKKEVHWLRKQTAMVFQQYHLFAHKTVIENVMEGLTIARKMRKQDAYAVA 120

Query: 121 RDELAKVGLADREKYYPRHLSGGQKQRVALARALAMKPDVLLLDEPTSALDPELVGEVEK 180
            +EL KVGL D+   YP  LSGGQKQRV +ARALA+ PDVLL DEPT+ALDPELVGEV +
Sbjct: 121 ENELRKVGLQDKLNAYPSQLSGGQKQRVGIARALAIHPDVLLFDEPTAALDPELVGEVLE 180

Query: 181 SIADAAKQGQTMVLVSHDMNFVYQVADKVLFLEKGRILESGTPEQLFNHPLEERTKEFFA 240
            + +  K G TM++V+H+M F  +V+D+V+F+++G I+E GTPE++F H  ++RT++F
Sbjct: 181 VMLEIVKTGATMIVVTHEMEFARRVSDQVVFMDEGVIVEQGTPEEVFRHTKKDRTRQFLR 240

Query: 241 SYNKSYL 247
             +  YL
Sbjct: 241 RVSPEYL 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2929> which encodes the amino acid sequence <SEQ ID 2930>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 199/247 (80%), Positives = 229/247 (92%)
```

```
Query: 1   MIKLRQLTKSFSGQKVLDKLDLDIEKGQVVALVGASGAGKSTFLRSMNYLEEPDYGTIEI 60
           MI +R L+K+FSGQKVLD L LDIEKGQV+ALVGASGAGKSTFLRS+NYLE+PD G+I I
Sbjct: 2   MITIRNLSKTFSGQKVLDSLALDIEKGQVIALVGASGAGKSTFLRSLNYLEKPDSGSISI 61

Query: 61  DDFKVDFKSISKDDILTLRRKLAMVFQQFNLFERRTALDNVKEGLKIVKKMSDQEATRIA 120
            DF VDF++I+ + +L LRRKLAMVFQQFNLFERRTAL+NVKEGLK+VKK+SDQEAT++A
Sbjct: 62  GDFTVDFETITTEQVLILRRKLAMVFQQFNLFERRTALENVKEGLKVVKKLSDQEATKLA 121

Query: 121 RDELAKVGLADREKYYPRHLSGGQKQRVALARALAMKPDVLLLDEPTSALDPELVGEVEK 180
           + ELAKVGLADR+ +YPRHLSGGQKQRVALARALAMKPDVLLLDEPTSALDPELVGEVEK
Sbjct: 122 QAELAKVGLADRKHHYPRHLSGGQKQRVALARALAMKPDVLLLDEPTSALDPELVGEVEK 181

Query: 181 SIADAAKQGQTMVLVSHDMNFVYQVADKVLFLEKGRILESGTPEQLFNHPLEERTKEFFA 240
           SI DAAK GQTMVLVSHDMNFVYQVAD+VLFL++G+ILE GTPE++F HP +ERTKEFFA
Sbjct: 182 SITDAAKSGQTMVLVSHDMNFVYQVADRVLFLDQGKILEQGTPEEVFRHPQKERTKEFFA 241

Query: 241 SYNKSYL 247
           SY+K+Y+
Sbjct: 242 SYSKTYI 248
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 961

A DNA sequence (GBSx1019) was identified in *S.agalactiae* <SEQ ID 2931> which encodes the amino acid sequence <SEQ ID 2932>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0831(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07290 GB:AP001519 thioredoxin reductase (NADPH) [Bacillus halodurans]
 Identities = 173/302 (57%), Positives = 234/302 (77%)

Query: 1   MYDTLIIGSGPGGMTAALYAARSNLKVGLIEQGAPGGQMNNTAEIENYPGYDHISGPELS 60
           +YD +I G+GP GMTAA+Y +R+NL   ++E+G PGGQM NT ++ENYPG+DHI GPELS
Sbjct: 7   VYDVVIAGAGPAGMTAAVYTSRANLSTVMVERGVPGGQMANTEDVENYPGFDHILGPELS 66

Query: 61  MKMYEPLEKFEVEHIYGIVQRVENDGDVKRVITEDESYEAKTVILATGAKNSLLGVPGEE 120
            KM+E  +KF  E+ YG ++ + + GD+K V   ++ Y+A+ VI+ATGA+   LGVPGE+
Sbjct: 67  TKMFEHAKKFGAEYAYGDIKEIIDQGDLKLVKAGNKEYKARAVIVATGAEYKKLGVPGEK 126

Query: 121 EYTSRGVSYCAVCDGAFFRDQDLLVVGGGDSAVEEAVFLTQFAKSVTIIHRRDQLRAQKV 180
           E + RGVSYCAVCDGAFF+ ++L+VVGGGDSAVEEAV+LT+FA  VTIIHRRDQLRAQK+
Sbjct: 127 ELSGRGVSYCAVCDGAFFKGKELVVVGGGDSAVEEAVYLTRFASKVTIIHRRDQLRAQKI 186

Query: 181 LQDRAFANEKIKFVWDSVVKEIKGNEIKVSGVTVENLKTGEISEMTFGGVFIYVGLKPHS 240
           LQ RAF N+KI+F+WD VVK+I G + KVS VT+E+ KTGE  +    GVFIY+G+ P +
Sbjct: 187 LQQRAFDNDKIEFIWDHVVKQINGTDGKVSSVTIEHAKTGEQQDFKTDGVFIYIGMLPLN 246

Query: 241 SMVSELGITDETGWVLTDTNMKTSIPGLYAIGDVRQKDLRQIATAVGEGAIAGQGVYNYI 300
             V  L I ++ G+++T+  M+TS+PG++A GDVR+K LRQI TA G+G++A Q V +YI
Sbjct: 247 EAVKNLNILNDEGYIVTNEEMETSVPGIFAAGDVREKSLRQIVTATGDGSLAAQNVQHYI 306

Query: 301 TE 302
            E
Sbjct: 307 EE 308
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2933> which encodes the amino acid sequence <SEQ ID 2934>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0386(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 236/300 (78%), Positives = 273/300 (90%)

Query: 1   MYDTLIIGSGPGGMTAALYAARSNLKVGLIEQGAPGGQMNNTAEIENYPGYDHISGPELS 60
           MYDTLIIGSGP GMTAALYAARSNL V +IEQGAPGGQMNNT +IENYPGYDHISGPEL+
Sbjct: 1   MYDTLIIGSGPAGMTAALYAARSNLSVAIIEQGAPGGQMNNTFDIENYPGYDHISGPELA 60

Query: 61  MKMYEPLEKFEVEHIYGIVQRVENDGDVKRVITEDESYEAKTVILATGAKNSLLGVPGEE 120
           MKMYEPLEKF VE+IYGIVQ++EN GD K V+TED SYEAKTVI+ATGAK  +LGVPGEE
Sbjct: 61  MKMYEPLEKFNVENIYGIVQKIENFGDYKCVLTEDASYEAKTVIIATGAKYRVLGVPGEE 120

Query: 121 EYTSRGVSYCAVCDGAFFRDQDLLVVGGGDSAVEEAVFLTQFAKSVTIIHRRDQLRAQKV 180
            YTSRGVSYCAVCDGAFFRDQDLLVVGGGDSAVEEA++LTQFAK VT++HRRDQLRAQK+
Sbjct: 121 YYTSRGVSYCAVCDGAFFRDQDLLVVGGGDSAVEEAIYLTQFAKKVTVVHRRDQLRAQKI 180

Query: 181 LQDRAFANEKIKFVWDSVVKEIKGNEIKVSGVTVENLKTGEISEMTFGGVFIYVGLKPHS 240
           LQDRAFAN+K+ F+WDSVVKEI+GN+IKVS V +EN+KTG++++  FGGVFIYVG+ P +
Sbjct: 181 LQDRAFANDKVDFIWDSVVKEIQGNDIKVSNVLIENVKTGQVTDHAFGGVFIYVGMNPVT 240

Query: 241 SMVSELGITDETGWVLTDTNMKTSIPGLYAIGDVRQKDLRQIATAVGEGAIAGQGVYNYI 300
            MV +L ITD  GW++TD +M+TSIPG++AIGDVRQKDLRQI TAVG+GAIAGQGVY+Y+
Sbjct: 241 GMVKDLEITDSEGWIITDDHMRTSIPGIFAIGDVRQKDLRQITTAVGDGAIAGQGVYHYL 300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 962

A DNA sequence (GBSx1020) was identified in *S.agalactiae* <SEQ ID 2935> which encodes the amino acid sequence <SEQ ID 2936>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3626(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15163 GB:Z99120 similar to nicotinate
           phosphoribosyltransferase [Bacillus subtilis]
 Identities = 309/476 (64%), Positives = 384/476 (79%), Gaps = 2/476 (0%)

Query: 2    YKDDSLTLHTDLYQINMMQVYFNKGIHNKRAVFEAYFRKVPFENGYAVFAGLERIVRYLE 61
            +KDDSL+LHTDLYQINM + Y+  GIH K+A+FE +FR++PFENGYAVFAGLE+ + YLE
Sbjct: 6    FKDDSLSLHTDLYQINMAETYWRDGIHEKKAIFELFFRRLPFENGYAVFAGLEKAIEYLE 65

Query: 62   NLSFSDSDLSYLE-ELGYPEEFLDYLKNLKMELTVKSAKEGDLVFANEPLVQIEGPLAQC 120
            N  F+DSDLSYL+ ELGY E+F++YL+ L    ++ S KEG+LVF NEP++++E PL +
Sbjct: 66   NFKFTDSDLSYLQDELGYHEDFIEYLRGLSFTGSLYSMKEGELVFNNEPIMRVEAPLVEA 125
```

```
Query: 121 QLVETAILNIINYQTLVATKAARIRSVIEDEPLLEFGTRRAQEMDAAIWGTRAAIIGGAN 180
           QL+ETA+LNI+NYQTL+ATKAARI+ VI DE  LEFGTRRA EMDAA+WG RAA+IGG +
Sbjct: 126 QLIETALLNIVNYQTLIATKAARIKGVIGDEVALEFGTRRAHEMDAAMWGARAALIGGFS 185

Query: 181 ATSNVRAGKIFNIPVSGTHAHALVQTYGDDYQAFKAYAETHKDCVFLVDTYDTLRVGVPN 240
           ATSNVRAGK FNIPVSGTHAHALVQ Y D+Y AFK YAETHKDCVFLVDTYDTLR G+PN
Sbjct: 186 ATSNVRAGKRFNIPVSGTHAHALVQAYRDEYTAFKKYAETHKDCVFLVDTYDTLRSGMPN 245

Query: 241 AIRVAKEMGEKINFLGVRLDSGDLAYLSKKVRQQLDDAGFPNAKIYASNDLDENTILNLK 300
           AIRVAKE G++INF+G+RLDSGDLAYLSKK R+ LD+AGF +AK+ AS+DLDE+TI+NLK
Sbjct: 246 AIRVAKEFGDRINFIGIRLDSGDLAYLSKKARKMLDEAGFTDAKVIASSDLDEHTIMNLK 305

Query: 301 MQKAKIDVWGVGTKLITAYDQPALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKKQ 360
            Q A+IDVWGVGTKLITAYDQPALGAVYK+V+IE D G M DTIK+S+N EKV+TPG+K+
Sbjct: 306 AQGARIDVWGVGTKLITAYDQPALGAVYKLVAIEED-GKMVDTIKISSNPEKVTTPGRKK 364

Query: 361 VWRITSRAKGKSEGDYITFADTDVTQLDEIEMFHPTYTYINKTVRDFDAVPLLVDIFDKG 420
           V+RI +++   SEGDYI   D  V      + MFHP +T+I+K V +F A  L   IF+KG
Sbjct: 365 VYRIINQSNHHSEGDYIALYDEQVNDQKRLRMFHPVHTFISKFVTNFYAKDLHELIFEKG 424

Query: 421 KLVYQLPSLQEIQEYGRKEFDQLWDEYKRVLNPQDYPVDLARDVWQNKMDLIDRIR 476
            L YQ P + +IQ+Y +     LW+EYKR+  P++YPVDL+ D W NKM  I  ++
Sbjct: 425 ILCYQNPEISDIQQYVQDNLSLLWEEYKRISKPEEYPVDLSEDCWSNKMQRIHEVK 480
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2937> which encodes the amino acid sequence <SEQ ID 2938>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3192(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 409/484 (84%), Positives = 446/484 (91%)

Query: 1   MYKDDSLTLHTDLYQINMMQVYFNKGIHNKRAVFEAYFRKVPFENGYAVFAGLERIVRYL 60
           MYKDDSLTLHTDLYQINMMQVYF +GIHN+ AVFE YFRK PF NGYAVFAGL+R+V YL
Sbjct: 1   MYKDDSLTLHTDLYQINMMQVYFEQGIHNRHAVFEVYFRKEPFNNGYAVFAGLQRMVEYL 60

Query: 61  ENLSFSDSDLSYLEELGYPEEFLDYLKNLKMELTVKSAKEGDLVFANEPLVQIEGPLAQC 120
           E   FS++DL+YLEELGYPE FL YLK L++ELT++SAKEGDLVFANEP+VQ+EGPL QC
Sbjct: 61  EQFQFSETDLAYLEELGYPENFLTYLKELRLELTIRSAKEGDLVFANEPIVQVEGPLGQC 120

Query: 121 QLVETAILNIINYQTLVATKAARIRSVIEDEPLLEFGTRRAQEMDAAIWGTRAAIIGGAN 180
           QLVETA+LNI+N+QTL+ATKAARIRSVIEDEPLLEFGTRRAQE+DAAIWGTRAA+IGGA+
Sbjct: 121 QLVETALLNIVNFQTLIATKAARIRSVIEDEPLLEFGTRRAQELDAAIWGTRAAMIGGAD 180

Query: 181 ATSNVRAGKIFNIPVSGTHAHALVQTYGDDYQAFKAYAETHKDCVFLVDTYDTLRVGVPN 240
           ATSNVRAGK F+IPVSGTHAHALVQ YG+DY AF AYA+THKDCVFLVDTYDTL+VGVP
Sbjct: 181 ATSNVRAGKRFDIPVSGTHAHALVQAYGNDYDAFMAYAKTHKDCVFLVDTYDTLKVGVPT 240

Query: 241 AIRVAKEMGEKINFLGVRLDSGDLAYLSKKVRQQLDDAGFPNAKIYASNDLDENTILNLK 300
           AIRVAKEMG+KINFLGVRLDSGDLAYLSK VRQQLDDAGF  AKIYASNDLDENTILNLK
Sbjct: 241 AIRVAKEMGDKINFLGVRLDSGDLAYLSKTVRQQLDDAGFTEAKIYASNDLDENTILNLK 300

Query: 301 MQKAKIDVWGVGTKLITAYDQPALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKKQ 360
           MQKAKIDVWGVGTKLITAYDQPALGAVYKIVSIE + GSMRDTIKLSNNAEKVSTPGKKQ
Sbjct: 301 MQKAKIDVWGVGTKLITAYDQPALGAVYKIVSIEQEDGSMRDTIKLSNNAEKVSTPGKKQ 360

Query: 361 VWRITSRAKGKSEGDYITFADTDVTQLDEIEMFHPTYTYINKTVRDFDAVPLLVDIFDKG 420
           VWRITSR KGKSEGDYITF D +V +L EIEMFHPTYTYI KTV++FDA+PLLVDIF KG
Sbjct: 361 VWRITSREKGKSEGDYITFTDINVNELTEIEMFHPTYTYIKKTVKEFDAIPLLVDIFVKG 420
```

```
Query: 421 KLVYQLPSLQEIQEYGRKEFDQLWDEYKRVLNPQDYPVDLARDVWQNKMDLIDRIRKEAL 480
           +LVYQLP+L EI+ Y +KEFD+LWDEYKRVLNPQDYPVDLARDVWQNKM LID IRK+A
Sbjct: 421 ELVYQLPTLAEIKAYAKKEFDKLWDEYKRVLNPQDYPVDLARDVWQNKMALIDNIRKDAY 480

Query: 481 AKGE 484
            K E
Sbjct: 481 GKSE 484
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 963

A DNA sequence (GBSx1021) was identified in *S.agalactiae* <SEQ ID 2939> which encodes the amino acid sequence <SEQ ID 2940>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2744(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC74810 GB:AE000269 NAD synthetase, prefers NH3 over glutamine
          [Escherichia coli K12]
 Identities = 173/274 (63%), Positives = 214/274 (77%), Gaps = 1/274 (0%)

Query: 1   MTLQDQIIKELGVKPVINPSQEIRRSVEFLKDYLLKHSFLKTYVLGISGGQDSTLAGRLA 60
           MTLQ QIIK LG KP IN  +EIRRSV+FLK YL  + F+K+ VLGISGGQDSTLAG+L
Sbjct: 1   MTLQQQIIKALGAKPQINAEEEIRRSVDFLKSYLQTYPFIKSLVLGISGGQDSTLAGKLC 60

Query: 61  QLAVEELRADTG-ENYQFIAIRLPYGIQADEEDAQKALDFIKPDIALTINIKEAVDGQVR 119
           Q+A+ ELR +TG E+ QFIA+RLPYG+QADE+D Q A+ FI+PD  LT+NIK AV    +
Sbjct: 61  QMAINELRLETGNESLQFIAVRLPYGVQADEQDCQDAIAFIQPDRVLTVNIKGAVLASEQ 120

Query: 120 ALNAAGVEITDFNKGNIKARQRMISQYAVAGQYAGAVIGTDHAAENITGFFTKFGDGGAD 179
           AL  AG+E++DF +GN KAR+RM +QY++AG  +G V+GTDHAAE ITGFFTK+GDGG D
Sbjct: 121 ALREAGIELSDFVRGNEKARERMKAQYSIAGMTSGVVVGTDHAAEAITGFFTKYGDGGTD 180

Query: 180 LLPLFRLNKSQGKQLLAELGADKALYEKIPTADLEENKPGIADEIALGVTYQEIDAYLEG 239
           + PL+RLNK QGKQLLA L   + LY+K PTADLE+++P + DE+ALGVTY  ID YLEG
Sbjct: 181 INPLYRLNKRQGKQLLAALACPEHLYKKAPTADLEDDRPSLPDEVALGVTYDNIDDYLEG 240

Query: 240 KVVSDKSRGIIENWWYKGQHKRHLPITIFDDFWK 273
           K V  +    IENW+ K +HKR  PIT+FDDFWK
Sbjct: 241 KNVPQQVARTIENWYLKTEHKRRPPITVFDDFWK 274
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2941> which encodes the amino acid sequence <SEQ ID 2942>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3482(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 213/274 (77%), Positives = 242/274 (87%), Gaps = 1/274 (0%)
```

```
Query: 1   MTLQDQIIKELGVKPVINPSQEIRRSVEFLKDYLLKHSFLKTYVLGISGGQDSTLAGRLA 60
           MTLQ++II++LGVK  I+P +EIR++V+FLK YL KHSFLKTYVLGISGGQDSTLAG+LA
Sbjct: 15  MTLQEEIIRQLGVKASIDPQEEIRKAVDFLKAYLRKHSFLKTYVLGISGGQDSTLAGKLA 74

Query: 61  QLAVEELRADTGEN-YQFIAIRLPYGIQADEEDAQKALDFIKPDIALTINIKEAVDGQVR 119
           Q+A+ ELR +  +  YQFIA+RLPYG+QADE DAQKAL FI PD  LTINIK AVDGQV
Sbjct: 75  QMAIAELREEASDQAYQFIAVRLPYGVQADEADAQKALAFIAPDQTLTINIKAAVDGQVE 134

Query: 120 ALNAAGVEITDFNKGNIKARQRMISQYAVAGQYAGAVIGTDHAAENITGFFTKFGDGGAD 179
           AL AAGVEI+DFNKGNIKARQRMISQYA+AGQ AGAVIGTDHAAENITGFFTKFGDGGAD
Sbjct: 135 ALQAAGVEISDFNKGNIKARQRMISQYAIAGQMAGAVIGTDHAAENITGFFTKFGDGGAD 194

Query: 180 LLPLFRLNKSQGKQLLAELGADKALYEKIPTADLEENKPGIADEIALGVTYQEIDAYLEG 239
           +LPLFRLNK QGK LL  LGAD ALYEK+PTADLE+ KPG+ADE+ALGVTYQ+ID YLEG
Sbjct: 195 ILPLFRLNKRQGKALLKVLGADAALYEKVPTADLEDQKPGLADEVALGVTYQDIDDYLEG 254

Query: 240 KVVSDKSRGIIENWWYKGQHKRHLPITIFDDFWK 273
           K++S  ++  IE WW+KGQHKRHLPITIFDDFWK
Sbjct: 255 KLISKVAQATIEKWWHKGQHKRHLPITIFDDFWK 288
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 964

A DNA sequence (GBSx1022) was identified in *S.agalactiae* <SEQ ID 2943> which encodes the amino acid sequence <SEQ ID 2944>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA82960 GB:Z30315 aminopeptidase C [Streptococcus thermophilus]
 Identities = 363/444 (81%), Positives = 407/444 (90%)

Query: 1   MSKLTQTFTDKLFADYQANTKFSAIENAVTHNGLLKSLETRQSEIENDYVFSIDLTKDEV 60
           M+ L+  FT+KLFADY+AN K+ AIENAVTHNGLLKS+ETRQSE+END+VFSIDLTKDEV
Sbjct: 1   MTSLSTDFTEKLFADYEANAKYGAIENAVTHNGLLKSIETRQSEVENDFVFSIDLTKDEV 60

Query: 61  SNQKQSGRCWMFAALNTFRHKLISDFKLENFELSQAHTFFWDKYEKSNWFMEQIIATANQ 120
           SNQK SGRCWMFAALNTFRHKLISDFKLE+FELSQAHTFFWDKYEKSNWF+EQIIATA+Q
Sbjct: 61  SNQKASGRCWMFAALNTFRHKLISDFKLESFELSQAHTFFWDKYEKSNWFLEQIIATADQ 120

Query: 121 ELSSRKVKFLLDVPQQDGGQWDMVVALFEKYGVVPKTVYPESVSSSASRELNQYLNKLLR 180
           E+ SRKVKFLLD PQQDGGQWDMVV+LFEKYGVVPK+VYPESV+SS SRELNQYLNKLLR
Sbjct: 121 EIGSRKVKFLLDTPQQDGGQWDMVVSLFEKYGVVPKSVYPESVASSNSRELNQYLNKLLR 180

Query: 181 QDAQILRELIAQGADGATVQNKKEELLQEIFNFLAMNLGLPPQSFDFAYRDKDNHYQSDK 240
           QDAQILR+LIA GAD A VQ KKEE LQEIFN+LAM LGLPP+ FDFAYRDKD++Y+S+K
Sbjct: 181 QDAQILRDLIASGADQAAVQAKKEEFLQEIFNYLAMTLGLPPRQFDFAYRDKDDNYRSEK 240

Query: 241 NITPKAFYQKYVNLDLSDYVSIINAPTVDKPYGQSYTVEMLGNVVGGPAVKYLNLDMKRF 300
            ITP+AF++KYV L LSDYVS+INAPT DKPYG+SYTVEMLGNVVG P+V+Y+NL M RF
Sbjct: 241 GITPRAFFEKYVGLKLSDYVSVINAPTADKPYGKSYTVEMLGNVVGAPSVRYINLPMDRF 300

Query: 301 KELAIAQMKSGETVWFGSDVGQVSNRQKGILATTTYDFNSSMDIKLSQDKAGRLDYSESL 360
           KELAIAQMK+GE+VWFGSDVGQVS+RQKGILAT  YDF +SMDI  +QDKAGRLDYSESL
Sbjct: 301 KELAIAQMKAGESVWFGSDVGQVSDRQKGILATNVYDFTASMDINWTQDKAGRLDYSESL 360

Query: 361 MTHAMVLTGVDLDESGQPLKWKVENSWGEKVGKDGYFVASDAWMDEYTYQIVVRKELLTK 420
```

```
           MTHAMVLTGVDLD  G+P+KWK+ENSWG+KVG+ GYFVASDAWMDEYTYQIVVRK+ LT
Sbjct: 361 MTHAMVLTGVDLDADGKPIKWKIENSWGDKVGQKGYFVASDAWMDEYTYQIVVRKDFLTA 420

Query: 421 EELEAYNAEPITLAPWDPMGALAN 444
           EEL AY A+P  LAPWDPMG+LA+
Sbjct: 421 EELAAYEADPQVLAPWDPMGSLAS 444
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2945> which encodes the amino acid sequence <SEQ ID 2946>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3002(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 369/443 (83%), Positives = 407/443 (91%)

Query: 1   MSKLTQTFTDKLFADYQANTKFSAIENAVTHNGLLKSLETRQSEIENDYVFSIDLTKDEV 60
           MS LT+TFT++LFA Y+AN KFSAIENAVTHNGLLKSLETRQSE++ND+VFSIDLTKD+V
Sbjct: 1   MSALTETFTEQLFAHYEANAKFSAIENAVTHNGLLKSLETRQSEVDNDFVFSIDLTKDKV 60

Query: 61  SNQKQSGRCWMFAALNTFRHKLISDFKLENFELSQAHTFFWDKYEKSNWFMEQIIATANQ 120
           SNQK SGRCWMFAALNTFRHKLI++FKLENFELSQAHTFFWDKYEK+NWFMEQ+IATA+Q
Sbjct: 61  SNQKASGRCWMFAALNTFRHKLITEFKLENFELSQAHTFFWDKYEKANWFMEQVIATADQ 120

Query: 121 ELSSRKVKFLLDVPQQDGGQWDMVVALFEKYGVVPKTVYPESVSSSASRELNQYLNKLLR 180
           EL+SRKVKFLLDVPQQDGGQWDMVV+LFEKYGVVPK+VYPES+SSS SRELNQYLNKLLR
Sbjct: 121 ELTSRKVKFLLDVPQQDGGQWDMVVSLFEKYGVVPKSVYPESISSSNSRELNQYLNKLLR 180

Query: 181 QDAQILRELIAQGADGATVQNKKEELLQEIFNFLAMNLGLPPQSFDFAYRDKDNHYQSDK 240
           QDAQILR+LIA GA    V+++K ELLQEIFNFLAM LGLPP+ FDFAYRDKD+HY  +K
Sbjct: 181 QDAQILRDLIASGAKADQVEDRKAELLQEIFNFLAMTLGLPPRHFDFAYRDKDDHYHVEK 240

Query: 241 NITPKAFYQKYVNLDLSDYVSIINAPTVDKPYGQSYTVEMLGNVVGGPAVKYLNLDMKRF 300
            +TP+AFY K+V L LSDYVS+INAPT DKPYG+SYTVEMLGNVVG   V+YLNLDMKRF
Sbjct: 241 GLTPQAFYDKFVGLKLSDYVSVINAPTADKPYGKSYTVEMLGNVVGSREVRYLNLDMKRF 300

Query: 301 KELAIAQMKSGETVWFGSDVGQVSNRQKGILATTTYDFNSSMDIKLSQDKAGRLDYSESL 360
           KELAI QM++GE+VWFGSDVGQVS+RQKGILAT TYDF +SMDI LSQDKAGRLDYSESL
Sbjct: 301 KELAIKQMQAGESVWFGSDVGQVSDRQKGILATNTYDFEASMDINLSQDKAGRLDYSESL 360

Query: 361 MTHAMVLTGVDLDESGQPLKWKVENSWGEKVGKDGYFVASDAWMDEYTYQIVVRKELLTK 420
           MTHAMVLTGVDLDE+G+PLKWKVENSWGEKVG  GYFVASDAWMDEYTYQIVVRKE LT
Sbjct: 361 MTHAMVLTGVDLDETGKPLKWKVENSWGEKVGDKGYFVASDAWMDEYTYQIVVRKEFLTA 420

Query: 421 EELEAYNAEPITLAPWDPMGALA 443
           +EL AY  EP  LAPWDPMGALA
Sbjct: 421 DELAAYEKEPQVLAPWDPMGALA 443
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 965

A DNA sequence (GBSx1024) was identified in *S.agalactiae* <SEQ ID 2947> which encodes the amino acid sequence <SEQ ID 2948>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have a cleavable N-term signal seq.
```

```
----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9533> which encodes amino acid sequence <SEQ ID 9534> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF17262 GB:AF210752 penicillin-binding protein 1A
          [Streptococcus pneumoniae]
 Identities = 412/725 (56%), Positives = 544/725 (74%), Gaps = 14/725 (1%)

Query: 4   IKKESVIKLLKYAFGIIMGFIILAIVIGGLLFAYYVSRSPKLTDQALKSVNSSLVYDGNN 63
           + K ++++L+KY     +  +I AIV+GG +F YYVS++P L++  L +  SS +YD  N
Sbjct: 1   MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESKLVATTSSKIYDNKN 60

Query: 64  KLIADLGSEKRESVSADSIPLNLVNAITSIEDKRFFKHRGVDIYRILGAAWHNLVSSNTQ 123
           +LIADLGSE+R +  A+ IP +LV AI SIED RFF HRG+D  RILGA   NL S++ Q
Sbjct: 61  QLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNSLQ 120

Query: 124 GGSTLDQQLIKLAYFSTNKSDQTLKRKSQEVWLALQMERKYTKEEILTFYINKVYMGNGN 183
           GGSTL QQLIKL YFST+ SDQT+ RK+QE WLA+Q+E+K TK+EILT+YINKVYM NGN
Sbjct: 121 GGSTLTQQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEILTYYINKVYMSNGN 180

Query: 184 YGMRTTAKSYFGKDLKELSIAQLALLAGIPQAPTQYDPYKNPESAQTRRNTVLQQMYQDK 243
           YGM+T A++Y+GKDL  LS+ QLALLAG+PQAP QYDPY +PE+AQ RRN VL +M
Sbjct: 181 YGMQTAAQNYYGKDLNNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQG 240

Query: 244 NISKKEYDQAVATPVTDGLKELKQKSTYPKYMDNYLKQVISEVKQKTGKDIFTAGLKVYT 303
            IS ++Y++AV TP+TDGL+ LK  S YP YMDNYLK+VI++V+++TG ++ T G+ VYT
Sbjct: 241 YISAEQYEKAVNTPITDGLQSLKSASNYPAYMDNYLKEVINQVEEETGYNLLTTGMDVYT 300

Query: 304 NINTDAQKQLYDIYNSDTYIAYPNNELQIASTIMDATNGKVIAQLGGRHQNENISFGTNQ 363
           N++ +AQK L+DIYN+D Y+AYP++ELQ+ASTI+D +NGKVIAQLG RHQ+ N+SFG NQ
Sbjct: 301 NVDQEAQKHLWDIYNTDEYVAYPDDELQVASTIVDVSNGKVIAQLGARHQSSNVSFGINQ 360

Query: 364 SVLTDRDWGSTMKPISAYAPAIDSGVYNSTGQSLNDSVYYWPGTSTQLYDWDRQYMGWMS 423
           +V T+RDWGSTMKPI+ YAPA++ GVY+ST   ++D  Y +PGT T +Y+WDR Y G ++
Sbjct: 361 AVETNRDWGSTMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTDTPVYNWDRGYFGNIT 420

Query: 424 MQTAIQQSRNVPAVRALEAAGLDEAKSFLEKLGIYYPEMNYSNAISSNNSSSDAKYGASS 483
           +Q A+QQSRNVPAV  L   GL+ AK+FL  LGI YP ++YSNAISSN + SD KYGASS
Sbjct: 421 LQYALQQSRNVPAVETLNKVGLNRAKTFLNGLGIDYPSLHYSNAISSNTTESDKKYGASS 480

Query: 484 EKMAAAYSAFANGGTYYKPQYVNKIEFSDGTNDTYAASGSRAMKETTAYMMTDMLKTVLT 543
           EKMAAAY+AFANGGTYYKP Y++K+ FSDG+   ++  G+RAMKETTAYMMTDM+KTVL
Sbjct: 481 EKMAAAYAAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLV 540

Query: 544 FGTGTKAAIPGVAQAGKTGTSNYTEDELAKIEATTGIYNSAVGTMAPDENFVGYTSKYTM 603
           +G G  A +P + QAGKTGTSNYT++E+ K       Y    G +APDE FVGYT KY M
Sbjct: 541 YGIGRGAYLPWLPQAGKTGTSNYTDEEIEK-------YIKNTGYVAPDEMFVGYTRKYAM 593

Query: 604 AIWTGYKNRLTPLYGSQLDIATEVYRAMMSYLTGGYSA-DWTMPEGLYRSGSYLYINGTT 662
           A+WTGY NRLTPL G  L +A +VYR+MM+YL+ G +  DW +PEGLYR+G +++ NG
Sbjct: 594 AVWTGYSNRLTPLVGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFVFKNGAR 653

Query: 663 TTGTYSSSVYKNIYQNSGQSSQSSSSTSSEKQKEDKNTANDANSSSPQVETPNNGNATTP 722
           +T  +SS   +      S +SS SSS +S+ +      + N++ +++P   T +    TTP
Sbjct: 654 ST--WSSPAPQQ--PPSTESSSSSSDSSTSQSNSTTPSTNNSTTTNPNNNTQQSN--TTP 707

Query: 723 NNSNQ 727
           +  NQ
Sbjct: 708 DQQNQ 712
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2949> which encodes the amino acid sequence <SEQ ID 2950>. Analysis of this protein sequence reveals the following:

```
     Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.96   Transmembrane   19 -  35 (   9 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA88918 GB:Z49095 penicillin-binding protein 1a [Streptococcus pneumoniae]
 Identities = 422/712 (59%), Positives = 536/712 (75%), Gaps = 8/712 (1%)

Query: 4   IKNPKILKWLKYVLSAILSLIILVIIIGGLLFTFYISSAPKLSEAQLKSTNSSLVYDGNN 63
           +  P IL+ +KY+  + LSL+I  I++GG +F +Y+S AP LSE++L +T SS +YD  N
Sbjct: 1   MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESKLVATTSSKIYDNKN 60

Query: 64  NLIADLGSEKRENVTADSIPINLVNAITSIEDKRFFNHRGVDLYRIFGAAFHNLTSQTTQ 123
            LIADLGSE+R N  A+ IP +LV AI SIED RFF+HRG+D  RI GA   NL S + Q
Sbjct: 61  QLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNSLQ 120

Query: 124 GGSTLDQQLIKLAYFSTNESDQTLKRKAQEVWLALQMERKYTKQEILTFYINKVYMGNGN 183
           GGSTL QQLIKL YFST+ SDQT+ RKAQE WLA+Q+E+K TKQEILT+YINKVYM NGN
Sbjct: 121 GGSTLTQQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEILTYYINKVYMSNGN 180

Query: 184 YGMLTAAKSYYGKDLKDLSYAQLALLAGIPQAPSQYDPYLHPEAAQNRRNVVLQQMYMEK 243
           YGM TAA++YYGKDL +LS  QLALLAG+PQAP+QYDPY HPEAAQ+RRN+VL +M  +
Sbjct: 181 YGMQTAAQNYYGKDLNNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQG 240

Query: 244 HLTKAEYETAIATPVAEGLQSLQQRSTYPKYMDNYLKQVIEEVKKETNKDIFTAGLKVYT 303
           +++  +YE A+ TP+ +GLQSL+  S YP YMDNYLK+VI +V++ET  ++ T G+ VYT
Sbjct: 241 YISAEQYEKAVNTPITDGLQSLKSASNYPAYMDNYLKEVINQVEEETGYNLLTTGMDVYT 300

Query: 304 NIIPDAQQTLYNIYHSGDYVYYPDQDFQVASTIVDVTNGHVIAQLGGRNQDENVSFGTNQ 363
           N+  +AQ+ L++IY+S  YV YPD D QVAST+VDV+NG VIAQLG R+Q  NVSFGTNQ
Sbjct: 301 NVDQEAQKHLWDIYNSDQYVSYPDDDLQVASTVVDVSNGKVIAQLGARHQASNVSFGTNQ 360

Query: 364 AVLTDRDWGSTMKPITAYAPAIESGVYTSTAQSTNDSVYYWPGTTTQLFNWDLRYNGWMT 423
           AV T+RDWGS+MKPIT YAPA+E GVY STA    +D  Y +PGT T L+NWD  Y G +T
Sbjct: 361 AVETNRDWGSSMKPITDYAPALEYGVYDSTASIVHDVPYNYPGTDTPLYNWDHVYFGNIT 420

Query: 424 IQAAIMLSRNVPAVRALEAAGLDYARSFLSSLGINYPEMHYSNAISSNNSSSDKKYGASS 483
           IQ A+  SRNV AV  L   GLD A++FL+ LGI+YP MHY+NAISSN + S+KKYGASS
Sbjct: 421 IQYALQQSRNVTAVETLNKVGLDRAKTFLNGLGIDYPSMHYANAISSNTTESNKKYGASS 480

Query: 484 EKMAAAYAAFANGGIYHKPRYVNKVEFSDGTSKTFDEKGKRAMKETTAYMMTDMLKTVLT 543
           EKMAAAYAAFANGGIYHKP Y+NK+ FSDG+ K F + G RAMKETTAYMMT+M+KTVLT
Sbjct: 481 EKMAAAYAAFANGGIYHKPMYINKIVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTVLT 540

Query: 544 YGTGTAAAIPGVAQAGKTGTSNYTDEELAKIGEKYGLYPDYVGTLAPDENFVGFTKRYAM 603
           YGTG  A +P + QAGKTGTSNYTDEE+ K       Y    G +APDE FVG+T++YAM
Sbjct: 541 YGTGRGAYLPWLPQAGKTGTSNYTDEEIEK-------YIKNTGYVAPDEMFVGYTRKYAM 593

Query: 604 AVWTGYKNRLTPVYGSSLEIASDVYRSMMTYLT-NGYSEDWTMPNGLYRSGGFLYLSGTY 662
           AVWTGY NRLTP+ G    +A  VYRSM+TYL+ +    DWTMP+GLYR+G F++ +G
Sbjct: 594 AVWTGYSNRLTPIIGDGFLVAGKVYRSMITYLSEDDQPGDWTMPDGLYRNGEFVFKNGAR 653

Query: 663 ASNTDYTNSVYNNLYSNNTTTASSQTTSDDTSSSNDTSNSTNTDNNGSHPST 714
           ++ +        +  S+++++ SS + S+ T+ S + S +TN +NN    +T
Sbjct: 654 STWSSPAPQQPPSTESSSSSSDSSTSQSNSTTPSTNNSTTTNPNNNTQQSNT 705
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 521/729 (71%), Positives = 621/729 (84%), Gaps = 10/729 (1%)

Query: 1   MITIKKESVIKLLKYAFGIIMGFIILAIVIGGLLFAYYVSRSPKLTDQALKSVNSSLVYD 60
           +ITIK   ++K LKY    I+  IIL I+IGGLLF +Y+S +PKL++  LKS NSSLVYD
Sbjct: 1   VITIKNPKILKWLKYVLSAILSLIILVIIIGGLLFTFYISSAPKLSEAQLKSTNSSLVYD 60

Query: 61  GNNKLIADLGSEKRESVSADSIPLNLVNAITSIEDKRFFKHRGVDIYRILGAAWHNLVSS 120
           GNN LIADLGSEKRE+V+ADSIP+NLVNAITSIEDKRFF HRGVD+YRI GAA+HNL S
Sbjct: 61  GNNNLIADLGSEKRENVTADSIPINLVNAITSIEDKRFFNHRGVDLYRIFGAAFHNLTSQ 120

Query: 121 NTQGGSTLDQQLIKLAYFSTNKSDQTLKRKSQEVWLALQMERKYTKEEILTFYINKVYMG 180
            TQGGSTLDQQLIKLAYFSTN+SDQTLKRK+QEVWLALQMERKYTK+EILTFYINKVYMG
Sbjct: 121 TTQGGSTLDQQLIKLAYFSTNESDQTLKRKAQEVWLALQMERKYTKQEILTFYINKVYMG 180

Query: 181 NGNYGMRTTAKSYFGKDLKELSIAQLALLAGIPQAPTQYDPYKNPESAQTRRNTVLQQMY 240
           NGNYGM T AKSY+GKDLK+LS AQLALLAGIPQAP+QYDPY +PE+AQ RRN VLQQMY
Sbjct: 181 NGNYGMLTAAKSYYGKDLKDLSYAQLALLAGIPQAPSQYDPYLHPEAAQNRRNVVLQQMY 240

Query: 241 QDKNISKKEYDQAVATPVTDGLKELKQKSTYPKYMDNYLKQVISEVKQKTGKDIFTAGLK 300
            +K+++K EY+ A+ATPV +GL+ L+Q+STYPKYMDNYLKQVI EVK++T KDIFTAGLK
Sbjct: 241 MEKHLTKAEYETAIATPVAEGLQSLQQRSTYPKYMDNYLKQVIEEVKKETNKDIFTAGLK 300

Query: 301 VYTNINTDAQKQLYDIYNSDTYIAYPNNELQIASTIMDATNGKVIAQLGGRHQNENISFG 360
           VYTNI  DAQ+ LY+IY+S  Y+ YP+ + Q+ASTI+D TNG VIAQLGGR+Q+EN+SFG
Sbjct: 301 VYTNIIPDAQQTLYNIYHSGDYVYYPDQDFQVASTIVDVTNGHVIAQLGGRNQDENVSFG 360

Query: 361 TNQSVLTDRDWGSTMKPISAYAPAIDSGVYNSTGQSLNDSVYYWPGTSTQLYDWDRQYMG 420
           TNQ+VLTDRDWGSTMKPI+AYAPAI+SGVY ST QS NDSVYYWPGT+TQL++WD +Y G
Sbjct: 361 TNQAVLTDRDWGSTMKPITAYAPAIESGVYTSTAQSTNDSVYYWPGTTTQLFNWDLRYNG 420

Query: 421 WMSMQTAIQQSRNVPAVRALEAAGLDEAKSFLEKLGIYYPEMNYSNAISSNNSSSDAKYG 480
           WM++Q AI  SRNVPAVRALEAAGLD A+SFL  LGI YPEM+YSNAISSNNSSSD KYG
Sbjct: 421 WMTIQAAIMLSRNVPAVRALEAAGLDYARSFLSSLGINYPEMHYSNAISSNNSSSDKKYG 480

Query: 481 ASSEKMAAAYSAFANGGTYYKPQYVNKIEFSDGTNDTYAASGSRAMKETTAYMMTDMLKT 540
           ASSEKMAAAY+AFANGG Y+KP+YVNK+EFSDGT+ T+   G RAMKETTAYMMTDMLKT
Sbjct: 481 ASSEKMAAAYAAFANGGIYHKPRYVNKVEFSDGTSKTFDEKGKRAMKETTAYMMTDMLKT 540

Query: 541 VLTFGTGTKAAIPGVAQAGKTGTSNYTEDELAKIEATTGIYNSAVGTMAPDENFVGYTSK 600
           VLT+GTGT AAIPGVAQAGKTGTSNYT++ELAKI    G+Y   VGT+APDENFVG+T +
Sbjct: 541 VLTYGTGTAAAIPGVAQAGKTGTSNYTDEELAKIGEKYGLYPDYVGTLAPDENFVGFTKR 600

Query: 601 YTMAIWTGYKNRLTPLYGSQLDIATEVYRAMMSYLTGGYSADWTMPEGLYRSGSYLYING 660
           Y MA+WTGYKNRLTP+YGS L+IA++VYR+MM+YLT GYS DWTMP GLYRSG +LY++G
Sbjct: 601 YAMAVWTGYKNRLTPVYGSSLEIASDVYRSMMTYLTNGYSEDWTMPNGLYRSGGFLYLSG 660

Query: 661 TTTTGT-YSSSVYKNIYQNSGQSSQSSSSTSSEKQKEDKNTANDANSSSPQVETPNNGNA 719
           T  + T Y++SVY N+Y N      ++++ SS+   +D +++ND ++S+    T NNG+
Sbjct: 661 TYASNTDYTNSVYNNLYSN------NTTTASSQTTSDDTSSSNDTSNST---NTDNNGSH 711

Query: 720 TTPNNSNQT 728
            + ++   T
Sbjct: 712 PSTDDKKTT 720
```

A related GBS gene <SEQ ID 8695> and protein <SEQ ID 8696> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      6.55
GvH: Signal Score (-7.5): -1.98
     Possible site: 36
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   4.03 threshold:  0.0
   PERIPHERAL  Likelihood =  4.03     201
 modified ALOM score:  -1.31

*** Reasoning Step: 3
```

```
----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.5/76.2% over 712aa
                                                            Streptococcus
pneumoniae
  GP|6563351| penicillin-binding protein 1A Insert characterized

ORF00399(310 - 2484 of 2850)
GP|6563351|gb|AAF17262.1|AF210752_1|AF210752(1 - 713 of 719) penicillin-binding protein 1A
{Streptococcus pneumoniae}
%Match = 43.8
%Identity = 57.5  %Similarity = 76.2
Matches = 412  Mismatches = 166  Conservative Sub.s = 134

237       267       297       327       357       387       417       447
LIISEKMDFS*RRVPFLKSLT*ILLKKNY*AVITIKKESVIKLLKYAFGIIMGFIILAIVIGGLLFAYYVSRSPKLTDQA
                                  : | ::::|:||     : ::| |||:|| :| ||||::| |::
                                  MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESK
                                          10        20        30        40

477       507       537       567       597       627       657       687
LKSVNSSLVYDGNNKLIADLGSEKRESVSADSIPLNLVNAITSIEDKRFFKHRGVDIYRILGAAWHNLVSSNTQGGSTLD
| :  || :||  |:||||||||:| :  |: || :|| || |||| ||| |||:|  |||||   || |:: ||||||
LVATTSSKIYDNKNQLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNSLQGGSTLT
            60        70        80        90       100       110       120

717       747       777       807       837       867       897       927
QQLIKLAYFSTNKSDQTLKRKSQEVWLALQMERKYTKEEILTFYINKVYMGNGNYGMRTTAKSYFGKDLKELSIAQLALL
|||||| ||||: ||||: ||:|| |||:|:|:| ||:||||:||||||| ||||||:| |::|:||||  ||: |||||
QQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEILTYYINKVYMSNGNYGMQTAAQNYYGKDLNNLSLPQLALL
           140       150       160       170       180       190       200

957       987       1017      1047      1077      1107      1137      1167
AGIPQAPTQYDPYKNPESAQTRRNTVLQQMYQDKNISKKEYDQAVATPVTDGLKELKQKSTYPKYMDNYLKQVISEVKQK
||:|||| ||||| :||:|| ||| || :|     || ::|::|| ||:||||: ||  | || |||||||:||::|:::
AGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQGYISAEQYEKAVNTPITDGLQSLKSASNYPAYMDNYLKEVINQVEEE
           220       230       240       250       260       270       280

1197      1227      1257      1287      1317      1347      1377      1407
TGKDIFTAGLKVYTNINTDAQKQLYDIYNSDTYIAYPNNELQIASTIMDATNGKVIAQLGGRHQNENISFGTNQSVLTDR
|| :::| |: ||||:: :|||:|:||||:| |:|||::|||:||||:| :||||||||| |||: |:||| ||:| |:|
TGYNLLTTGMDVYTNVDQEAQKHLWDIYNTDEYVAYPDDELQVASTIVDVSNGKVIAQLGARHQSSNVSFGINQAVETNR
           300       310       320       330       340       350       360

1437      1467      1497      1527      1557      1587      1617      1647
DWGSTMKPISAYAPAIDSGVYNSTGQSLNDSVYYWPGTSTQLYDWDRQYMGWMSMQTAIQQSRNVPAVRALEAAGLDEAK
|||||||||: ||||:: |||:||   ::|  | :||| | :|:||| | | :::| |:|||||||||  |   ||: ||
DWGSTMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTDTPVYNWDRGYFGNITLQYALQQSRNVPAVETLNKVGLNRAK
           380       390       400       410       420       430       440

1677      1707      1737      1767      1797      1827      1857      1887
SFLEKLGIYYPEMNYSNAISSNNSSSDAKYGASSEKMAAAYSAFANGGTYYKPQYVNKIEFSDGTNDTYAASGSRAMKET
:||  ||| || ::|||||||| : || |||||||||||||:||||||||||| |::|: ||||:   ::  |:||||||
TFLNGLGIDYPSLHYSNAISSNTTESDKKYGASSEKMAAAYAAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKET
           460       470       480       490       500       510       520

1917      1947      1977      2007      2037      2067      2097      2127
TAYMMTDMLKTVLTFGTGTKAAIPGVAQAGKTGTSNYTEDELAKIEATTGIYNSAVGTMAPDENFVGYTSKYTMAIWTGY
||||||||:|||| :| |  | :| : |||||||||||::|: |       |    | :|||| ||||| || ||:||||
TAYMMTDMMKTVLVYGIGRGAYLPWLPQAGKTGTSNYTDEEIEK-------YIKNTGYVAPDEMFVGYTRKYAMAVWTGY
           540       550       560       570              580       590
```

```
2157      2187      2214      2244      2274      2304      2334      2364
KNRLTPLYGSQLDIATEVYRAMMSYLT-GGYSADWTMPEGLYRSGSYLYINGTTTTGTYSSSVYKNIYQNSGQSSQSSSS
 |||||| |  | :| :|||:||:||: |    || :||||||:| ::: ||  :|  :||    :    | :|| |||
SNRLTPLVGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFVFKNGARST--WSSPAPQQ--PPSTESSSSSSD
        610       620       630       640       650         660         670

2394      2424      2454      2484      2514      2544      2574      2604
TSSEKQKEDKNTANDANSSSPQVETPNNGNATTPNNSNQTVPGTGHGNGNGNGNNNTVPNGN*KTGYIIQFFNL*LLLLI
:|: :      : |:: :::|   |  :   |||:  ||
SSTSQSNSTTPSTNNSTTTNPNNNTQQS--NTTPDQQNQNPQPAQP
            690       700         710
```

SEQ ID 8696 (GBS146) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 4; MW 82kDa), in Figure 168 (lane 11-13; MW 96.5kDa) and in Figure 238 (lane 8; MW 96.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 2; MW 107kDa).

Purified Thio-GBS146-His is shown in Figure 244, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 966

A DNA sequence (GBSx1025) was identified in *S.agalactiae* <SEQ ID 2951> which encodes the amino acid sequence <SEQ ID 2952>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3647(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA26957 GB:M90528 ORF [Streptococcus oralis]
 Identities = 143/196 (72%), Positives = 165/196 (83%), Gaps = 1/196 (0%)

Query: 1   MVNYPHQLIRKTTVTKSKKKKIDFANRGMSFEAAINATNDYYLSHELAVIHKKPTPVQIV 60
            MVNYPH++  +       + K +FANRGMSFE  INATNDYYLSH LAVIHKKPTP+QIV
Sbjct: 1   MVNYPHKISSQKRQAPPSQTK-NFANRGMSFEKMINATNDYYLSHGLAVIHKKPTPIQIV 59

Query: 61  KVDYPKRSRAKIVEAYFRQASTTDYSGVYKGYYIDFEAKETRQKTAMPMKNFHAHQIEHM 120
            +VDYP+RSRAKIVEAYFRQASTTDYSGVY GYYIDFEAKETRQK A+PMKNFH HQI+HM
Sbjct: 60  RVDYPQRSRAKIVEAYFRQASTTDYSGVYDGYYIDFEAKETRQKHAIPMKNFHHHQIQHM 119

Query: 121 ANVLQQKGICFVLLHFSTLKETYLLPANELISFYQIDKGNKSMPIDYIRKNGFFVKESAF 180
              VL Q+GICFVLLHF++ +ETYLLPA +LI FY  DKG KSMP+ YIR+NG+ ++  AF
Sbjct: 120 EQVLAQRGICFVLLHFASQQETYLLPAVDLIRFYHQDKGQKSMPLGYIRENGYRIELGAF 179

Query: 181 PQVPYLDIIEEKLLGG 196
            PQ+PYLDII+E LLGG
Sbjct: 180 PQIPYLDIIKEHLLGG 195
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2953> which encodes the amino acid sequence <SEQ ID 2954>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
          bacterial cytoplasm --- Certainty=0.5030(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 166/199 (83%), Positives = 177/199 (88%)

Query: 1   MVNYPHQLIRKTTVTKSKKKKIDFANRGMSFEAAINATNDYYLSHELAVIHKKPTPVQIV 60
           MVNYPH LIR+   +  K+ K+DFANRGMSFEAAINATNDYYLS ++AVIHKKPTPVQIV
Sbjct: 1   MVNYPHNLIRQKVSSVQKQNKVDFANRGMSFEAAINATNDYYLSRQIAVIHKKPTPVQIV 60

Query: 61  KVDYPKRSRAKIVEAYFRQASTTDYSGVYKGYYIDFEAKETRQKTAMPMKNFHAHQIEHM 120
           KVDYPKRSRAKIVEAYFRQASTTDY GVYKG+Y+DFEAKETRQKTAMPMKNFH HQIEHM
Sbjct: 61  KVDYPKRSRAKIVEAYFRQASTTDYCGVYKGHYVDFEAKETRQKTAMPMKNFHLHQIEHM 120

Query: 121 ANVLQQKGICFVLLHFSTLKETYLLPANELISFYQIDKGNKSMPIDYIRKNGFFVKESAF 180
           A VL QKGICFVLLHFSTLKETY LPA  LISFYQID G+KSMPIDYIRKNGF V   AF
Sbjct: 121 ACVLHQKGICFVLLHFSTLKETYYLPAQALISFYQIDNGSKSMPIDYIRKNGFKVAFGAF 180

Query: 181 PQVPYLDIIEEKLLGGDYN 199
           PQVPYL+IIE+  LGGDYN
Sbjct: 181 PQVPYLNIIEQNFLGGDYN 199
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 967

A DNA sequence (GBSx1026) was identified in *S.agalactiae* <SEQ ID 2955> which encodes the amino acid sequence <SEQ ID 2956>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3227(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14136 GB:Z99115 similar to hypothetical proteins from B. subtilis [Bacillus
subtilis]
 Identities = 74/174 (42%), Positives = 97/174 (55%), Gaps = 6/174 (3%)

Query: 5   ILVTGYKNFELGIFQDKDPRITIIKKAIDKDFRRFLENGADWFIFMGNLGFEYWALEVAL 64
           + +TGYK FELGIF+  D  +  IKKAI      FL+ G +W +  G LG E WA E A
Sbjct: 4   LAITGYKPFELGIFKQDDKALYYIKKAIKNRLIAFLDEGLEWILISGQLGVELWAAEAAY 63

Query: 65  DLQKEY-DFQIATIFTFENHGQNWNEANKAKL-ALFKQVDF-VKYTFPSYENPGQFKQYN 121
           DLQ+EY D ++A I  F    +NW E NK +  A+  Q D+    T   YE+P QFKQ N
Sbjct: 64  DLQEEYPDLKVAVITPFYEQEKNWKEPNKEQYEAVLAQADYEASLTHRPYESPLQFKQKN 123

Query: 122 HFLINNTQGAYLFYDSENETNLKFLLEMMEKK---EAYDISFLTFDRLNEIYEE 172
            F I+ + G  L YD E E + K++L   EK+   + Y I F+T D L    EE
Sbjct: 124 QFFIDKSDGLLLLYDPEKEGSPKYMLGTAEKRREQDGYPIYFITMDDLRVTVEE 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2957> which encodes the amino acid sequence <SEQ ID 2958>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                 bacterial cytoplasm --- Certainty=0.3041(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 102/167 (61%), Positives = 127/167 (75%)

Query: 3   STILVTGYKNFELGIFQDKDPRITIIKKAIDKDFRRFLENGADWFIFMGNLGFEYWALEV 62
           + IL+TGY++FE+GIF  KDPR++IIK+AI KD   +LENG DWFIF GNLGFE WALEV
Sbjct: 2   TAILITGYRSFEIGIFDHKDPRVSIIKQAIRKDLIGYLENGVDWFIFTGNLGFEQWALEV 61

Query: 63  ALDLQKEYDFQIATIFTFENHGQNWNEANKAKLALFKQVDFVKYTFPSYENPGQFKQYNH 122
           A +L++EY  QIATIF FE HG  WNE NK  L+ F+ VDFVKY FP+YE P QF QY
Sbjct: 62  ANELKEEYPLQIATIFLFETHGDRWNEKNKEVLSQFRAVDFVKYYFPNYEQPTQFSQYYQ 121

Query: 123 FLINNTQGAYLFYDSENETNLKFLLEMMEKKEAYDISFLTFDRLNEI 169
           FL+  T+GAY+FYD+ENETNLK+ L+  +    Y +  LTFDRLN++
Sbjct: 122 FLLEKTEGAYVFYDTENETNLKYFLKKAKDMPHYQLLLLTFDRLNDM 168
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 968

A DNA sequence (GBSx1027) was identified in *S.agalactiae* <SEQ ID 2959> which encodes the amino acid sequence <SEQ ID 2960>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.5188(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 969

A DNA sequence (GBSx1028) was identified in *S.agalactiae* <SEQ ID 2961> which encodes the amino acid sequence <SEQ ID 2962>. This protein is predicted to be cell division protein DivIVA. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.2736(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9535> which encodes amino acid sequence <SEQ ID 9536> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14135 GB:Z99115 ypsB [Bacillus subtilis]
 Identities = 46/102 (45%), Positives = 69/102 (67%), Gaps = 14/102 (13%)

Query: 14  SPKDIFEQDFKVSMRGYDKKEVDVFLDDVIKDYENYLEQIEKLQMENRRLQQALDKKESE 73
            S K+I E++FK  +RGY +++VD FLD +IKDYE + ++IE+LQ EN +L++ L+    E
Sbjct: 9   SAKEILEKEFKTGVRGYKQEDVDKFLDMIIKDYETFHQEIEELQQENLQLKKQLE----E 64

Query: 74  ASNVRNSGTAMYNQKPIAQSATNFDILKRISRLEKEVFGRQI 115
            AS           ++P+  + TNFDILKR+S LEK VFG ++
Sbjct: 65  AS----------KKQPVQSNTTNFDILKRLSNLEKHVFGSKL 96
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2963> which encodes the amino acid sequence <SEQ ID 2964>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4466(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 71/112 (63%), Positives = 85/112 (75%), Gaps = 6/112 (5%)

Query: 8   MASIIYSPKDIFEQDFKVSMRGYDKKEVDVFLDDVIKDYENYLEQIEKLQMENRRLQQAL 67
           M SIIYSPKDIFEQ+FK SMRG+DKKEVD FLD+VIKDYEN+  QIE L+ EN    +AL
Sbjct: 1   MTSIIYSPKDIFEQEFKTSMRGFDKKEVDEFLDNVIKDYENFNAQIEALKAEN----EAL 56

Query: 68  DKKESEASNVRNSGTAMYNQKP--IAQSATNFDILKRISRLEKEVFGRQIRE 117
             K + +A N  ++       +P  +AQSATNFDILKRIS+LEKEVFG+QI E
Sbjct: 57  KKAKFQARNTVSATVQQPVPQPTRVAQSATNFDILKRISKLEKEVFGKQIIE 108
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 970

A DNA sequence (GBSx1029) was identified in *S.agalactiae* <SEQ ID 2965> which encodes the amino acid sequence <SEQ ID 2966>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0655(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14134 GB:Z99115 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 204/382 (53%), Positives = 274/382 (71%), Gaps = 3/382 (0%)

Query: 3   ESFKLIATAAAGLEAIVGREIRNLGIDCQVENGRVRFHGDIKTIIETNLWLRAADRIKII 62
           + + LIATA  G+EA+V +E+R+LG +C+V+NG+V F GD   I   NLWLR ADRIK+
Sbjct: 2   KKYTLIATAPMGIEAVVAKEVRDLGYECKVDNGKVIFEGDALAICRANLWLRTADRIKVQ 61

Query: 63  VGEFPAPTFEELFQGVYGLDWENYLPLGAKFPIAKAKCVKSKLHNEPSVQAISKKAVAKK 122
           V  F A TF+ELF+     ++W +++P   KFP+   K VKS L + P  Q I KKA+ +K
Sbjct: 62  VASFKAKTFDELFEKTKAINWRSFIPENGKFPVI-GKSVKSTLASVPDCQRIVKKAIVEK 120

Query: 123 LQKVFHRPEGVPLQENGAEFKIEVSILKDKATVMIDTTGSSLFKRGYRAEKGGAPIKENM 182
           L K+        ++E GAE+K+E+S+LKD+A + +D++G+ L KRGYR ++GGAPIKE +
```

```
Sbjct: 121 L-KLQSGKANDWIEETGAEYKVEISLLKDQALITLDSSGTGLHKRGYRVDQGGAPIKETL 179

Query: 183 AAAIIQLSNWFPDKPLIDPTCGSGTFCIEAAMIGMNIAPGFNRDFAFEAWPWVDQSQVQK 242
           AAA++QL+NW PD+P +DP CGSGT  IEAA+IG NIAPGFNRDF  E W W+ +    K
Sbjct: 180 AAALVQLTNWTPDRPFVDPFCGSGTIAIEAALIGQNIAPGFNRDFVSEDWEWIGKDLWNK 239

Query: 243 VRDEAESKANYDIDLDISGFDLDGRMVEIARKNAEEAGLGDVIKLKQMRLQDLKTDKING 302
            R E E KANYD  L I   D+D RMV+IA++NAEEAGLGD+I+ KQM+++D  T+   G
Sbjct: 240 ARLEVEEKANYDQPLTIFASDIDHRMVQIAKENAEEAGLGDLIQFKQMQVKDFTTNLEFG 299

Query: 303 VIISNPPYGERLLDDKAVDILYNEMGQTFAPLKTWSKFILTSDEGFEKKYGSQADKKRKL 362
           VI+ NPPYGERL + KAV+ +Y EMGQ F PL TWS ++LTS+E FE+ YG +A KKRKL
Sbjct: 300 VIVGNPPYGERLGEKKAVEQMYKEMGQAFEPLDTWSVYMLTSNENFEEAYGRKATKKRKL 359

Query: 363 YNGTLKVDLYQYYGERVRRQVK 384
           +NG +K D YQY+  +VR Q K
Sbjct: 360 FNGFIKTDYYQYW-SKVRPQRK 380
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2967> which encodes the amino acid sequence <SEQ ID 2968>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0324(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 317/383 (82%), Positives = 354/383 (91%)

Query: 1   MKESFKLIATAAAGLEAIVGREIRNLGIDCQVENGRVRFHGDIKTIIETNLWLRAADRIK 60
           MKE+F+L+ATAAAGLEA+VG+E+R LG DCQVENG+V F GD++ I++TNLWLRAADRIK
Sbjct: 1   MKETFRLVATAAAGLEAVVGKEVRALGFDCQVENGKVYFEGDVEAIVKTNLWLRAADRIK 60

Query: 61  IIVGEFPAPTFEELFQGVYGLDWENYLPLGAKFPIAKAKCVKSKLHNEPSVQAISKKAVA 120
           IIVG+FPA TFEELFQGV+ LDWENYLPLGAKFPI+KAKCVKSKLHNEPSVQAI+KKAV
Sbjct: 61  IIVGQFPARTFEELFQGVFALDWENYLPLGAKFPISKAKCVKSKLHNEPSVQAITKKAVV 120

Query: 121 KKLQKVFHRPEGVPLQENGAEFKIEVSILKDKATVMIDTTGSSLFKRGYRAEKGGAPIKE 180
           KKLQK FHRPEGVPLQE G+ F IEVSILKD+AT+MIDTTGSSLFKRGYR +KGGAPIKE
Sbjct: 121 KKLQKHFHRPEGVPLQEVGSTFNIEVSILKDQATIMIDTTGSSLFKRGYRVQKGGAPIKE 180

Query: 181 NMAAAIIQLSNWFPDKPLIDPTCGSGTFCIEAAMIGMNIAPGFNRDFAFEAWPWVDQSQV 240
           NMAAAI+ LSNWFPDKPL+DPTCGSGTFCIEAAMIGMNIAPGFNR FAFE W WVD+  V
Sbjct: 181 NMAAAILALSNWFPDKPLVDPTCGSGTFCIEAAMIGMNIAPGFNRSFAFEEWSWVDKDMV 240

Query: 241 QKVRDEAESKANYDIDLDISGFDLDGRMVEIARKNAEEAGLGDVIKLKQMRLQDLKTDKI 300
           Q+VRD+AE +ANY+I+LDISGFD+DGRM+EIA+ NAEEAGL DVI  KQMRLQD +TDK+
Sbjct: 241 QQVRDDAEQEANYEIELDISGFDIDGRMIEIAKSNAEEAGLSDVITFKQMRLQDFRTDKV 300

Query: 301 NGVIISNPPYGERLLDDKAVDILYNEMGQTFAPLKTWSKFILTSDEGFEKKYGSQADKKR 360
           NGV+ISNPPYGERLLDDKAVDILYNEMGQTFAPLKTWSKFILTSDE FE KYG +ADKKR
Sbjct: 301 NGVVISNPPYGERLLDDKAVDILYNEMGQTFAPLKTWSKFILTSDELFELKYGQKADKKR 360

Query: 361 KLYNGTLKVDLYQYYGERVRRQV 383
           KLYNGTLKVDLYQ+YGERV+R +
Sbjct: 361 KLYNGTLKVDLYQFYGERVKRHL 383
```

SEQ ID 2966 (GBS255) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 7; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 4; MW 69kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 971

A DNA sequence (GBSx1030) was identified in *S.agalactiae* <SEQ ID 2969> which encodes the amino acid sequence <SEQ ID 2970>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.02   Transmembrane  171 - 187 ( 167 - 193)

----- Final Results -----
               bacterial membrane --- Certainty=0.7007(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD16120 GB:AF094508 dentin phosphoryn [Homo sapiens]
 Identities = 71/398 (17%), Positives = 152/398 (37%), Gaps = 16/398 (4%)

Query: 16  TDGLEFKDAK-EMTVEEAVRKDSEIKAGITEEDSILDKYIKQHRDEVASQKFETKSSDFA 74
           +D  +  D+K + +  E+   DS+ K+  ++ +S          D   S      S
Sbjct: 152 SDSSDSSDSKSDSSKSESDSSDSDSKSDSSDSNSSDSSDNSDSSDSSNSSNSSDSSDSSD 211

Query: 75  NLDTASLDDFIKKQREELSAMLAAEELSKKLDNSVSQEQDTEANAVSPKEESSQEQENSV 134
           + D++S  D       + S    +   S+  D+S S + D+  ++ S    SS    ++
Sbjct: 212 SSDSSSSD--SSNSSDSSDSSDSSNSSESSDSSDSSDSDSSDSSDSSNSNSSDSDSSNS 269

Query: 135 TPVPPLNTEAEPTATEPDSTIADSEEYKSSSKKRGGIVGTLIALILLLIVAIFGYNYFKN 194
           +     +  ++ + +   S  +DS +   SS        +  +       +    N   +
Sbjct: 270 SDSSDSSNSSDSSDSSDSSNSSDSSDSSDSSNSSDSSDSSDSS------DSSDSSNSSDS 323

Query: 195 NNSTNSQTATSQSSSSKATTTSSEEDKKASQNLDNFNKSYANFFVDDKKTQLKNSEFDKL 254
           N+S+NS  ++  S SS ++ +S   D   S + D+ N S       D      +S+
Sbjct: 324 NDSSNSSDSSDSSDSSDSSNSSDSSDSSDSSDSDSSNSS-------DSSNSSDSSDSCNS 376

Query: 255 SELEKKVDALKGTKYYGKVKVKFDSLKRQIDAVKAVNDKFKSPAVVDGKKSEKLEVKDGA 314
           S+     D+  G+      +    +     D+  + N    S +    + S   +  D +
Sbjct: 377 SDSSDSSDSSDGSDSDSSNRSDSSNSSDSSDSSDSSNSSDSSDSSDSNESSNSSDSSDSS 436

Query: 315 NFDSLDSKTLNTGNASLDSLLHSIVSTGRNQVKQSEEQASSNKVSDTQITEQPNVTNGQS 374
           N    DS   +  + S DS   S  S   N    S    SSN    +  ++  N ++  +
Sbjct: 437 NSSDSDSSDSSNSSDSSDSSNSSDSSESSNSSDNSNSSDSSNSSDSSDSSDSSNSSDSSN 496

Query: 375 SSSAATINNQAAGTASGNLERNRSRVPYNNAAIADTGN 412
           SS ++  ++ +   +S + + + S    +++  +D+ +
Sbjct: 497 SSDSSNSSDSSDSNSSDSSDSSXSSDSSDSSDSSDSSD 534
 Identities = 64/341 (18%), Positives = 140/341 (40%), Gaps = 35/341 (10%)

Query: 59  DEVASQKFETKSSDFANLDTASLDDFIKKQREELS-AMLAAEELSKKLDNSVSQEQDTEA 117
           D+  S K ++ SSD  + D+++  D       + S +  +++  S   D+S S + D+
Sbjct: 76  DKSDSGKGKSDSSDSDSSDSSNSSDSSDSSDSDSSDSNSSSDSDSSDSDSSDSSDSDSSD 135

Query: 118 NAVSPKEESSQEQENSVTPVPPLNTEAEPTATEPDSTIADSEEYKSSSKKRGGIVGTLIA 177
           ++ S     S +  +S       +++++ + +E DS+ +DS+   S S
Sbjct: 136 SSNSSDSSDSSDSSDSSDSSDSSDSKSDSSKSESDSSDSDSKSDSSDSN----------- 184

Query: 178 LILLLIVAIFGYNYFKNNNSTNSQTATSQSSSSKATTTSSEEDKKASQNLDNFNKSYANF 237
                           +++S NS ++ S +SS+ + ++ S +   +S + D+ N S ++
Sbjct: 185 ---------------SSDSSDNSDSSDSSNSSNSSDSSDSSDSSDSSSSSDSSNSSDSS- 228

Query: 238 FVDDKKTQLKNSEFDKLSELEKKVDALKGTKYYGKVKVKFDSLKRQIDAVKAVNDKFKSP 297
              D      +SE    S+     D+   +          DS     D+  + N    S
```

```
Sbjct: 229 ---DSSDSSNSSESSDSSD-SSDSDSSDSSDSSNSNSSDSDS-SNSSDSSDSSNSSDSSD 283

Query: 298 AVVDGKKSEKLEVKDGANFDSLDSKTLNTGNASLDSLLHSIVSTGRNQVKQSEEQASSNK 357
           +      S+  +  D +N  S DS   +  + S DS   S  +   N    S+   SS+
Sbjct: 284 SSDSSNSSDSSDSSDSSN--SSDSSDSSDSSDSSDSSNSSDSNDSSNSSDSSDSSDSSDS 341

Query: 358 VSDTQITEQPNVTNGQSSSSAATINNQAAGTASGNLERNRS 398
            + +  ++  + ++  SS+S+ + N+  +  +  + + + S
Sbjct: 342 SNSSDSSDSSDSSDSDSSNSSDSSNSSDSSDSCNSSDSSDS 382
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2971> which encodes the amino acid sequence <SEQ ID 2972>. Analysis of this protein sequence reveals the following:

```
        Possible site: 28

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.70   Transmembrane  180 - 196 ( 175 - 202)

----- Final Results -----
               bacterial membrane --- Certainty=0.6880(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF15293 GB:AF202180 erythrocyte membrane-associated giant
           protein antigen 332 [Plasmodium falciparum]
 Identities = 41/173 (23%), Positives = 87/173 (49%), Gaps = 10/173 (5%)

Query: 1   VSEESKEVEVTKESQTLGLNEAKSMTIGEAVRKQSE----IKAGVTKDDSILDKYIKQHR 56
           + E  + V + KE +  GL+  + +   ++V +Q+E    I   + K+ S ++    ++
Sbjct: 78  IEEAEENVWIEKEVEEEGLDNEEVIDEEDSVSEQAEEEVYINEEILKESSDVEDVKVENE 137

Query: 57  ---DEVSSQKFDAKYTELDTASLDNFIKKQREALSKAGLVDDEPVSAESAEQDSTLVEEV 113
              +EV+ +         +   LDN++ ++ E++++  +VD+ P S E  E +S ++EE+
Sbjct: 138 LMNEEVNEETQSVAENNEEDKELDNYVVEETESVTEEVVVDEVPNSKEVQEIES-IIEEI 196

Query: 114 AEDLAPMETTAVVTGIPVEATVPVLDLDPSERVIPEPQMTKEEPKRDQFLSED 166
            ED    +      G  +E  V  +  D SE ++ E  +T+E  K++  ++ED
Sbjct: 197 VEDGLTTDDLVGQQGSVIEEVVEEVGSD-SEGIVEEASITEEVEKKES-VTED 247
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 234/506 (46%), Positives = 304/506 (59%), Gaps = 36/506 (7%)

Query: 1   MSEDQKHPFFEPKKETDGLEFKDAKEMTVEEAVRKDSEIKAGITEEDSILDKYIKQHRDE 60
           +SE+ K    E  KE+  L   +AK MT+ EAVRK SEIKAG+T++DSILDKYIKQHRDE
Sbjct: 1   VSEESKE--VEVTKESQTLGLNEAKSMTIGEAVRKQSEIKAGVTKDDSILDKYIKQHRDE 58

Query: 61  VASQKFETKSSDFANLDTASLDDFIKKQREELSAMLAAEELSKKLDNSVSQEQDTEANAV 120
           V+SQKF+ K   +  LDTASLD+FIKKQRE LS    A  +  +  ++ S EQD+
Sbjct: 59  VSSQKFDAK---YTELDTASLDNFIKKQREALSK---AGLVDDEPVSAESAEQDSTLVEE 112

Query: 121 SPKEESSQEQENSVTPVPPLNT--------------EAEPTATEP--DSTIADSEEYKSS 164
             ++ +  E    VT +P   T              E + T  EP  D  +++   + +
Sbjct: 113 VAEDLAPMETTAVVTGIPVEATVPVLDLDPSERVIPEPQMTKEEPKRDQFLSEDSHHPAK 172

Query: 165 SKKRGGIVGTLIALILLLIVAIFGYNYFKNNNSTNSQTATSQSSSSKATTTSSEEDKKAS 224
              + G +  L  L+L ++  +FG+N+F   +S  +    S+ + +   T S+++  +
Sbjct: 173 QNTKKGWLIALFLLLLAILAVVFGWNHFLRQDSGKTTQTASKQTKTSLQTDSAKKATRLK 232

Query: 225 QNLDNFNKSYANFFVDDKKTQLKNSEFDKLSELEKKVDALKGTKYYGKVKVKFDSLKRQI 284
                F K Y  F+ D  K++LKNS F  L +LE  + AL+G+ YY K K K DSLK+ I
Sbjct: 233 AAAKAFEKLYGTFYTDATKSKLKNSAFATLPDLEAALKALEGSAYYDKAKAKVDSLKKAI 292

Query: 285 DAVKAVNDKFKSPAVVDGKKSEKLEVKDGANFDSLDSKTLNTGNASLDSLLHSIVSTGRN 344
            A+ AVN KF S  VVDG+K    EVK  ANFD L S TL  GNA+LD++L + ++ GR
```

```
Sbjct: 293 AAITAVNGKFVSDVVVDGEKVSA-EVKADANFDDLSSATLTIGNANLDAVLQASITEGRQ 351

Query: 345 QVKQSEEQASSNKVSDTQITEQPNVTNGQSSSSAATINNQAAGTAS---GNLERNRSRVP 401
           Q+    E A   K ++ Q   Q     GQS+S A +      G  S    +L+R+ SRVP
Sbjct: 352 QLASKAEAA---KAANEQAV-QDQAAQGQSTSVAPS----GYGLTSYDPASLQRHLSRVP 403

Query: 402 YNNAAIADTGNPAWIFNPGVLEKIVATSQARGYFSGNNYILEPVNIINGNGYYNMFKLDG 461
           YN   IAD  NP+W FNPGVLEKIVATSQARGY SGN YILEPVNIINGNGYYNMFK DG
Sbjct: 404 YNQDVIADRANPSWAFNPGVLEKIVATSQARGYISGNQYILEPVNIINGNGYYNMFKPDG 463

Query: 462 TYLFSINAKTGYFVGNAPGRADSLDY 487
           TYLFSIN KTGYFVGN  G AD+LDY
Sbjct: 464 TYLFSINCKTGYFVGNGKGYADALDY 489
```

SEQ ID 2970 (GBS351) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 2; MW 57kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 5; MW 82kDa).

GBS351-GST was purified as shown in Figure 216, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 972

A DNA sequence (GBSx1031) was identified in *S.agalactiae* <SEQ ID 2973> which encodes the amino acid sequence <SEQ ID 2974>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3169(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2975> which encodes the amino acid sequence <SEQ ID 2976>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3169(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 129/160 (80%), Positives = 149/160 (92%)

Query: 1   MTKEVVVESFELDHTIVKAPYVRLISEEVGPVGDIITNFDIRLIQPNENAIDTAGLHTIE 60
           MTKEV+VESFELDHTIVKAPYVRLISEE GP GD ITNFD+RL+QPN+N+I+TAGLHTIE
Sbjct: 1   MTKEVIVESFELDHTIVKAPYVRLISEEFGPKGDRITNFDVRLVQPNQNSIETAGLHTIE 60

Query: 61  HLLAKLIRQRINGLIDCSPFGCRTGFHMIMWGKQDATEIAKVIKSSLEAIAGGVTWEDVP 120
           HLLAKLIRQRI+G+IDCSPFGCRTGFH+IMWGK  +T+IAKVIKSSLE IA G+TWEDVP
Sbjct: 61  HLLAKLIRQRIDGMIDCSPFGCRTGFHLIMWGKHSSTDIAKVIKSSLEEIATGITWEDVP 120

Query: 121 GTTIESCGNYKDHSLHSAQEWAKLILSQGISDNAFERHIV 160
           GTT+ESCGNYKDHSL +A+EWA+LI+ QGISD+ F RH++
Sbjct: 121 GTTLESCGNYKDHSLFAAKEWAQLIIDQGISDDPFSRHVI 160
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 973

A DNA sequence (GBSx1032) was identified in *S.agalactiae* <SEQ ID 2977> which encodes the amino acid sequence <SEQ ID 2978>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF34762 GB:AF228345 unknown [Listeria monocytogenes]
 Identities = 302/532 (56%), Positives = 400/532 (74%), Gaps = 14/532 (2%)

Query: 4    IILAMVCALIGLIIGYVAISMKMKSSKEAAELTLLNAEQDAVDLRGKAEIEAEHIRKAAE 63
            I + ++ +L+ LI+G V  S+  KSS           E+     RG AE+  E  +K AE
Sbjct: 3    IAITIISSLLFLIVGLVVGSLIFKSS----------TEKKLAAARGTAELIVEDAKKEAE 52

Query: 64   RESKAHQKELLLEAKEEARKYREEIEKEFKSDRQELKQMEARLTDRASSLDRKDENLSNK 123
                  +KE LLEAKEE  + R EIE E +  R E ++ E RL  R  +LDRKD +LS +
Sbjct: 53   TT----KKEALLEAKEENHRLRTEIENELRGRRTETQKAENRLLQREENLDRKDTSLSKR 108

Query: 124 EKMLDSKEQSLTDKSRHINEREQEIATLETKKVEELSRIAELSQEEAKDIILADTEKDLA 183
           E  L+ KE+S++ + + I E+E ++A +   +  EL RI+ LS+EEAK IIL   E++L
Sbjct: 109 EATLERKEESISKRQQQIEEKESKLAEMIQAEQTELERISALSKEEAKSIILNQVEEELT 168

Query: 184 HDIATRIKEAEREVKDRSNKIAKDLLAQAMQRLAGEYVTEQTITTVHLPDDNMKGRIIGR 243
           HD A  +KE+E   K+ S+K AK++L+ A+QR A ++V E T++ V LP+D MKGRIIGR
Sbjct: 169 HDTAIMVKESENRAKEESDKKAKNILSLAIQRCAADHVAETTVSVVTLPNDEMKGRIIGR 228

Query: 244 EGRNIRTLESLTGIDVIIDDTPEVVVLSGFDPIRREIARMTLESLIQDGRIHPARIEELV 303
           EGRNIRTLE+LTGID+IIDDTPE V+LSGFDPIRREIAR+ LE L+QDGRIHPARIEE+V
Sbjct: 229 EGRNIRTLETLTGIDLIIDDTPEAVILSGFDPIRREIARIALEKLVQDGRIHPARIEEMV 288

Query: 304 EKNRLEMDQRIREYGEAAAYEIGAPNLHPDLIKIMGRLQFRTSYGQNVLRHSVEVGKLAG 363
           +K R E+D+ IRE GE A +E+G  ++HPDLIKI+GRL++RTSYGQNVL HS+EV KLAG
Sbjct: 289 DKARKEVDEHIREVGEQATFEVGIHSIHPDLIKILGRLRYRTSYGQNVLNHSLEVSKLAG 348

Query: 364 ILAGELGENVDLARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPIVVNTIASHHG 423
           ILAGELGE+V LA+RAG LHD+GKAID E+EGSHVEIG+E A KYKE+ +V+N+IASHHG
Sbjct: 349 ILAGELGEDVTLAKRAGLLHDIGKAIDHEIEGSHVEIGVELATKYKENDVVINSIASHHG 408

Query: 424 DVEPDSVIAVIVAAADALSSARPGARNESMENYIKRLRDLEEIANGFEGVQNAFALQAGR 483
           D E  SVIAV+VAAADALS+ARPGAR+E++ENYI+RL  LEEI+  ++GV+ ++A+QAGR
Sbjct: 409 DTEATSVIAVLVAAADALSAARPGARSETLENYIRRLEKLEEISESYDGVEKSYAIQAGR 468

Query: 484 EIRIMVQPGKVSDDQVVIMSHKVREKIEQNLDYPGNIKVTVIREMRAVDFAK 535
           E+RI+V+P  + D     ++  +R++IE+ LDYPG+IKVTVIRE RAV++AK
Sbjct: 469 EVRIIVEPDTIDDLSSYRLARDIRKRIEEELDYPGHIKVTVIRETRAVEYAK 520
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2979> which encodes the amino acid sequence <SEQ ID 2980>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

```
  bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF34762 GB:AF228345 unknown [Listeria monocytogenes]
 Identities = 299/534 (55%), Positives = 408/534 (75%), Gaps = 14/534 (2%)

Query: 2   VNIILLIVSALIGLILGYALISIRLKSAKEAAELTLLNAEQEAVDIRGKAEVDAEHIKKT 61
           + I + I+S+L+ LI+G  + S+  KS+           E++    RG AE+    I +
Sbjct: 1   MTIAITIISSLLFLIVGLVVGSLIFKSS----------TEKKLAAARGTAEL----IVED 46

Query: 62  AKRESKANRKELLLEAKEEARKYREEIEQEFKSERQELKQLETRLAERSLTLDRKDENLS 121
           AK+E++  +KE LLEAKEE  + R EIE E +  R E ++ E RL +R   LDRKD +LS
Sbjct: 47  AKKEAETTKKEALLEAKEENHRLRTEIENELRGRRTETQKAENRLLQREENLDRKDTSLS 106

Query: 122 SKEKVLDSKEQSLTDKSKHIDERQLQVEKLEEEKKAELEKVAAMTIAEAREVILMETENK 181
            +E  L+ KE+S++ + + I+E++ ++ ++ + ++ ELE+++A++  EA+ +IL + E +
Sbjct: 107 KREATLERKEESISKRQQQIEEKESKLAEMIQAEQTELERISALSKEEAKSIILNQVEEE 166

Query: 182 LTHEIATRIRDAERDIKDRTVKTAKDLLAQAMQRLAGEYVTEQTITSVHLPDDNMKGRII 241
           LTH+ A  ++++E   K+ + K AK++L+ A+QR A ++V E T++ V LP+D MKGRII
Sbjct: 167 LTHDTAIMVKESENRAKEESDKKAKNILSLAIQRCAADHVAETTVSVVTLPNDEMKGRII 226

Query: 242 GREGRNIRTLESLTGIDVIIDDTPEVVILSGFDPIRREIARMTLESLIADGRIHPARIEE 301
           GREGRNIRTLE+LTGID+IIDDTPE VILSGFDPIRREIAR+ LE L+ DGRIHPARIEE
Sbjct: 227 GREGRNIRTLETLTGIDLIIDDTPEAVILSGFDPIRREIARIALEKLVQDGRIHPARIEE 286

Query: 302 LVEKNRLEMDNRIREYGEAAAYEIGAPNLHPDLIKIMGRLQFRTSFGQNVLRHSVEVGKL 361
           +V+K R E+D  IRE GE A +E+G  ++HPDLIKI+GRL++RTS+GQNVL HS+EV KL
Sbjct: 287 MVDKARKEVDEHIREVGEQATFEVGIHSIHPDLIKILGRLRYRTSYGQNVLNHSLEVSKL 346

Query: 362 AGILAGELGENVALARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPVVVNTIASH 421
           AGILAGELGE+V LA+RAG LHD+GKAID E+EGSHVEIG+E A KYKE+ VV+N+IASH
Sbjct: 347 AGILAGELGEDVTLAKRAGLLHDIGKAIDHEIEGSHVEIGVELATKYKENDVVINSIASH 406

Query: 422 HGDVEPDSVIAVLVAAADALSSARPGARNESMENYIKRLRDLEEIATSFDGVQNSFALQA 481
           HGD E  SVIAVLVAAADALS+ARPGAR+E++ENYI+RL  LEEI+ S+DGV+ S+A+QA
Sbjct: 407 HGDTEATSVIAVLVAAADALSAARPGARSETLENYIRRLEKLEEISESYDGVEKSYAIQA 466

Query: 482 GREIRIMVQPEKISDDQVVILSHKVREKIENNLDYPGNIKVTVIREMRAVDYAK 535
           GRE+RI+V+P+ I D     L+  +R++IE  LDYPG+IKVTVIRE RAV+YAK
Sbjct: 467 GREVRIIVEPDTIDDLSSYRLARDIRKRIEEELDYPGHIKVTVIRETRAVEYAK 520
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 451/535 (84%), Positives = 503/535 (93%)

Query: 1   MFNIILAMVCALIGLIIGYVAISMKMKSSKEAAELTLLNAEQDAVDLRGKAEIEAEHIRK 60
           M NIIL +V ALIGLI+GY  IS+++KS+KEAAELTLLNAEQ+AVD+RGKAE++AEHI+K
Sbjct: 1   MVNIILLIVSALIGLILGYALISIRLKSAKEAAELTLLNAEQEAVDIRGKAEVDAEHIKK 60

Query: 61  AAERESKAHQKELLLEAKEEARKYREEIEKEFKSDRQELKQMEARLTDRASSLDRKDENL 120
            A+RESKA++KELLLEAKEEARKYREEIE+EFKS+RQELKQ+E RL +R+ +LDRKDENL
Sbjct: 61  TAKRESKANRKELLLEAKEEARKYREEIEQEFKSERQELKQLETRLAERSLTLDRKDENL 120

Query: 121 SNKEKMLDSKEQSLTDKSRHINEREQEIATLETKKVEELSRIAELSQEEAKDIILADTEK 180
           S+KEK+LDSKEQSLTDKS+HI+ER+ ++  LE +K  EL ++A ++  EA+++IL +TE
Sbjct: 121 SSKEKVLDSKEQSLTDKSKHIDERQLQVEKLEEEKKAELEKVAAMTIAEAREVILMETEN 180

Query: 181 DLAHDIATRIKEAEREVKDRSNKIAKDLLAQAMQRLAGEYVTEQTITTVHLPDDNMKGRI 240
            L H+IATRI++AER++KDR+ K AKDLLAQAMQRLAGEYVTEQTIT+VHLPDDNMKGRI
Sbjct: 181 KLTHEIATRIRDAERDIKDRTVKTAKDLLAQAMQRLAGEYVTEQTITSVHLPDDNMKGRI 240

Query: 241 IGREGRNIRTLESLTGIDVIIDDTPEVVVLSGFDPIRREIARMTLESLIQDGRIHPARIE 300
           IGREGRNIRTLESLTGIDVIIDDTPEVV+LSGFDPIRREIARMTLESLI DGRIHPARIE
Sbjct: 241 IGREGRNIRTLESLTGIDVIIDDTPEVVILSGFDPIRREIARMTLESLIADGRIHPARIE 300
```

```
Query: 301 ELVEKNRLEMDQRIREYGEAAAYEIGAPNLHPDLIKIMGRLQFRTSYGQNVLRHSVEVGK 360
           ELVEKNRLEMD RIREYGEAAAYEIGAPNLHPDLIKIMGRLQFRTS+GQNVLRHSVEVGK
Sbjct: 301 ELVEKNRLEMDNRIREYGEAAAYEIGAPNLHPDLIKIMGRLQFRTSFGQNVLRHSVEVGK 360

Query: 361 LAGILAGELGENVDLARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPIVVNTIAS 420
           LAGILAGELGENV LARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHP+VVNTIAS
Sbjct: 361 LAGILAGELGENVALARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPVVVNTIAS 420

Query: 421 HHGDVEPDSVIAVIVAAADALSSARPGARNESMENYIKRLRDLEEIANGFEGVQNAFALQ 480
           HHGDVEPDSVIAV+VAAADALSSARPGARNESMENYIKRLRDLEEIA  F+GVQN+FALQ
Sbjct: 421 HHGDVEPDSVIAVLVAAADALSSARPGARNESMENYIKRLRDLEEIATSFDGVQNSFALQ 480

Query: 481 AGREIRIMVQPGKVSDDQVVIMSHKVREKIEQNLDYPGNIKVTVIREMRAVDFAK 535
           AGREIRIMVQP K+SDDQVVI+SHKVREKIE NLDYPGNIKVTVIREMRAVD+AK
Sbjct: 481 AGREIRIMVQPEKISDDQVVILSHKVREKIENNLDYPGNIKVTVIREMRAVDYAK 535
```

SEQ ID 2978 (GBS86) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 6; MW 59kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 5; MW 84kDa).

GBS86-GST was purified as shown in Figure 192, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 974

A DNA sequence (GBSx1033) was identified in *S.agalactiae* <SEQ ID 2981> which encodes the amino acid sequence <SEQ ID 2982>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 975

A DNA sequence (GBSx1034) was identified in *S.agalactiae* <SEQ ID 2983> which encodes the amino acid sequence <SEQ ID 2984>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.87   Transmembrane  146 - 162 ( 146 - 162)

----- Final Results -----
                bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8697> which encodes amino acid sequence <SEQ ID 8698> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -10.72
GvH: Signal Score (-7.5): -5.66
     Possible site: 29
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -2.87 threshold:  0.0
   INTEGRAL    Likelihood = -2.87   Transmembrane  138 - 154 ( 138 - 154)
   PERIPHERAL  Likelihood =  3.76      51
 modified ALOM score:   1.07

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG21390 GB:AF302051 ABC transporter ATP binding subunit
          [Bacillus licheniformis]
 Identities = 84/218 (38%), Positives = 138/218 (62%), Gaps = 1/218 (0%)

Query: 12  DIIKVDHIFKSIGQKTILEDISFSIASNQCVALIGPNGAGKTTLMSTLLGDISISSGSLT 71
            +++ + ++ K+  QKT ++ I FSI   + VA++GPNGAGKTT +S +LG +  ++G++T
Sbjct: 3   NVVSLTNVTKTFRQKTAVDQIDFSIKKGEIVAILGPNGAGKTTTISMILGLLKPTAGNIT 62

Query: 72  IFNLPAHHNRLKYKVAILPQE-NVLPSKFTVRELIDFQRCLFPEVLPMSLILDYLQWSDT 130
            +F+   H  R++ K+  + QE +V+P      E+I+  R  +P+ L    +      +D
Sbjct: 63  LFDSMPHEKRVREKIGTMLQEVSVMPGLRCRVEIIELIRSYYPKPLSFQKLRTLTGLTDK 122

Query: 131 HLQQFTETLSGGQKRLLAFVLTLVGKPQLLFLDEPTSGMDTSTRQRFWELIATLKKEGVT 190
             L+   E LSGGQKR L F L L G P+L+  DEPT GMD ++R RFW+ + +L ++G T
Sbjct: 123 DLKTQAEKLSGGQKRRLGFALALAGDPELMIFDEPTVGMDITSRNRFWQTVQSLAEQGKT 182

Query: 191 IVYSSHYIEEVEHTADRILVLHKGKLLRDTTPLCHEAR 228
            I++S+HY++E +  A RIL+   GK++ D TPL  ++R
Sbjct: 183 IIFSTHYLQEADDAAQRILLFKDGKIVADGTPLQIKSR 220
```

There is also homology to SEQ ID 686.

SEQ ID 8698 (GBS350) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 13; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 4; MW 54kDa).

GBS350-GST was purified as shown in Figure 226, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 976

A DNA sequence (GBSx1035) was identified in *S.agalactiae* <SEQ ID 2985> which encodes the amino acid sequence <SEQ ID 2986>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
```

```
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 977

A DNA sequence (GBSx1036) was identified in *S.agalactiae* <SEQ ID 2987> which encodes the amino acid sequence <SEQ ID 2988>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.51   Transmembrane  222 - 238 ( 214 - 241)
   INTEGRAL    Likelihood = -6.90   Transmembrane  104 - 120 ( 101 - 125)
   INTEGRAL    Likelihood = -5.84   Transmembrane  140 - 156 ( 138 - 159)
   INTEGRAL    Likelihood = -5.20   Transmembrane   19 -  35 (  18 -  41)
   INTEGRAL    Likelihood = -1.28   Transmembrane  164 - 180 ( 164 - 180)

----- Final Results -----
              bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB69806 GB:AJ243712 YVFS protein [Bacillus cereus]
 Identities = 73/239 (30%), Positives = 127/239 (52%), Gaps = 4/239 (1%)

Query: 9   KMEFLLTKRQLANLIMAIGMPVAFFLFFSGFMGEGLTKAIEAIYVRNYMITMAGFSSLSF 68
           K+E L T R    +  ++ MPV F+  F+  +           +  +Y+I+MA FS +
Sbjct: 4   KIEILRTFRNKLFIFFSLLMPVMFYYIFTNVVQ---VPQNGDAWKAHYLISMATFSIVGT 60

Query: 69  AFFTFPFSMKDDQLSNRMQLLRHSPVPMWQYYLAKIIRILFYYCLAITVVFLTGHILRQV 128
           A F+F   +  ++      LL+ +P+P   Y  AKII        +I V+F+ G ++  V
Sbjct: 61  ALFSFGVRLSQERGQGWTHLLKITPLPEGAYLTAKIIAQTVVNAFSILVIFIAGILINHV 120

Query: 129 SMPIEQWMQSFLLLLGGATCFIPFGLLVSYFKNTELMSMVANICYMSLAVLGGMWMPITM 188
            + I QW+ + L LL G T F+  G ++   K  +  + +ANI  MSLA++GG+WMPI +
Sbjct: 121 ELTIGQWIGAGLWLLLGVTPFLALGTVIGSIKKADAAAGLANILNMSLAIVGGLWMPIEV 180

Query: 189 FPKWLQALSKLTPTYHLTQVILSPFANSFAGF-SLIILIGYGIIMLVIAYLLSQKRHSI 246
           FPK L+ + + TPTYH         A    G+ ++ +L GY +I +V++  + +++ ++
Sbjct: 181 FPKILRTIGEWTPTYHFGSGAWDIVAGKSIGWENIAVLGGYFLIFVVVSIYIRKRQEAV 239
```

There is also homology to SEQ ID 682 and to SEQ ID 1628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 978

A DNA sequence (GBSx1037) was identified in *S.agalactiae* <SEQ ID 2989> which encodes the amino acid sequence <SEQ ID 2990>. This protein is predicted to be histidine kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.43   Transmembrane  105 - 121 ( 102 - 124)
```

```
   INTEGRAL    Likelihood = -6.95   Transmembrane  130 - 146 ( 129 - 149)

----- Final Results -----
              bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9537> which encodes amino acid sequence <SEQ ID 9538> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54584 GB:AJ006400 histidine kinase [Streptococcus pneumoniae]
 Identities = 138/350 (39%), Positives = 212/350 (60%), Gaps = 3/350 (0%)

Query: 11  MYFIPLVFLIYPIGGILYYHYPFWTLFFTLAFVGAYLYSVIIRGESKYHMIAWSTMLTYI 70
           M++I L+F+I+PI  ++      W L   + FV AYL  V+     +   + W  MLTY+
Sbjct: 11  MFWISLIFMIFPILSVVTGWLSAWHLLIDILFVVAYL-GVLTTKSQRLSWLYWGLMLTYV 69

Query: 71  FYMTIFINSGFIWYIYFLSNLLVYRFRDK-LKSFRFISFACTLATVVF-LCFFKASDFGD 128
              T F+   +IW+ +FLSNLL Y F  + LKS    +F      VV  L  F+  +
Sbjct: 70  VGNTAFVAVNYIWFFFFLSNLLSYHFSVRSLKSLHVWTFLLAQVLVVGQLLIFQRIEVEF 129

Query: 129 RIMFLIVPIFCIGYMWIAIENRNSEEQREKIAEQNQYINILSAENERNRIGRDLHDSLGH 188
               L++  F     +  +  R  E+ +E   +QN  IN+L AENER+RIG+DLHDSLGH
Sbjct: 130 LFYLLVILTFVDLMTFGLVRIRIVEDLKEAQVKQNAQINLLLAENERSRIGQDLHDSLGH 189

Query: 189 TFAMMTLKTELALKLLEKRNYDKVQKELSELNHISHQSMSEVRQIVSNLKYRTVVEEIDE 248
           TFAM+++KT+LAL+L +   Y +V+KEL E++ IS  SM+EVR IV NLK RT+  E++
Sbjct: 190 TFAMLSVKTDLALQLFQMEAYPQVEKELKEIHQISKDSMNEVRTIVENLKSRTLTSELET 249

Query: 249 LYRLFQLSNIKLTVVNKLETSQLSPVTQSTITMILKELSNNIVKHAEADSVELSLVRQGA 308
           + ++ +++ I++ V N L+ S L+   +ST +MIL EL  NI+KHA+A  V L L R
Sbjct: 250 VKKMLEIAGIEVQVENHLDKSSLTQELESTASMILLELVTNIIKHAKASKVYLKLERTEK 309

Query: 309 TINIEMIDNGCGFTNLDGDELHSIQERLTIVEGTLTILSRSKPTHIQVVL 358
            + + + D+GCGF ++ GDELH+++ R+    G ++++S+  PT +QV L
Sbjct: 310 ELILTVRDDGCGFASISGDELHTVRNRVFPFSGEVSVISQKHPTEVQVRL 359
```

There is also homology to SEQ ID 2992.

A related GBS gene <SEQ ID 8699> and protein <SEQ ID 8700> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      10.90
GvH: Signal Score (-7.5): -2.42
     Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 2 value:  -7.43 threshold:  0.0
   INTEGRAL    Likelihood = -7.43   Transmembrane  105 - 121 ( 102 - 124)
   INTEGRAL    Likelihood = -6.95   Transmembrane  130 - 146 ( 129 - 149)
   PERIPHERAL  Likelihood =  0.16      61
 modified ALOM score:   1.99

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 979

A DNA sequence (GBSx1038) was identified in *S.agalactiae* <SEQ ID 2993> which encodes the amino acid sequence <SEQ ID 2994>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane   49 -  65 (  49 -  65)

----- Final Results -----
              bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54585 GB:AJ006400 response regulator [Streptococcus pneumoniae]
 Identities = 95/153 (62%), Positives = 125/153 (81%), Gaps = 3/153 (1%)

Query: 1   MKLLVAEDQSMLRDAMCQLLLMEESVSTIDQAGNGGEAIAILSNKAIDVAILDVEMPILS 60
           MK+LVAEDQSMLRDAMCQLL+++  V ++ QA NG EAI +L  +++D+AILDVEMP+ +
Sbjct: 1   MKVLVAEDQSMLRDAMCQLLMLQPDVESVFQAKNGQEAIQLLEKESVDIAILDVEMPVKT 60

Query: 61  GLDVLEWVRKYQ-NVKVIIVTTFKRSGYFQRAIRSNVDAYVLKDRSVADLMKTIQKVLSG 119
           GL+VLEW+R  +   KV++VTTFKR GYF+RA+++ VDAYVLK+R++ADLM+T+  VL G
Sbjct: 61  GLEVLEWIRAEKLETKVVVVTTFKRPGYFERAVKAGVDAYVLKERNIADLMQTLHTVLEG 120

Query: 120 GKEYSPELMENVI--SNPLSEQEIKILSLIAQG 150
            KEYSPELME V+   NPL+EQEI +L  IAQG
Sbjct: 121 RKEYSPELMEVVMMHPNPLTEQEIAVLKGIAQG 153
```

There is also homology to SEQ ID 2996.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 980

A DNA sequence (GBSx1039) was identified in *S.agalactiae* <SEQ ID 2997> which encodes the amino acid sequence <SEQ ID 2998>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.69   Transmembrane  158 - 174 ( 145 - 184)
   INTEGRAL    Likelihood = -4.94   Transmembrane   11 -  27 (   8 -  31)
   INTEGRAL    Likelihood = -3.93   Transmembrane   74 -  90 (  73 -  92)
   INTEGRAL    Likelihood = -2.39   Transmembrane  103 - 119 ( 102 - 119)
   INTEGRAL    Likelihood = -1.86   Transmembrane   42 -  58 (  38 -  59)

----- Final Results -----
              bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB85965 GB:AE000909 unknown [Methanothermobacter
           thermoautotrophicus]
 Identities = 46/183 (25%), Positives = 81/183 (44%), Gaps = 11/183 (6%)

Query: 5   KERFDTLSDAILAIAMTILVLEI-------KTPATMGDIGDFTRNIGLFIVSFVVVFNFW 57
           K+R + L DAI AIAMTILVL I         PA    I     ++  + +SF+++  FW
Sbjct: 6   KKRLEGLVDAIFAIAMTILVLGIDVPTGTMSVPAMDAYIMGLASDLYSYCLSFLLLGVFW 65
```

```
Query: 58   YERAQNSLDAQKTNDEIIALDIIEHLGICLIPLFTKFMISFENHNFAVMAYGLLTLLVGL 117
            +    +    +K +   I ++I+  + + L+P  TK   ++ +     + + L  L +GL
Sbjct: 66   WVNHMHFEKLEKVDTGFIWINIVWLMVVVLVPFSTKLTGNYGDLVTPNILFHLNMLTIGL 125

Query: 118 TSDIIRIRLASYDLVTIPSELKERVIKVMTTFAIRSVVVRFIIIILAYFLPEVGIFAYLV 177
               +  I      L+ I       ++K      + ++    + +IL    PE    AY V
Sbjct: 126 LLSMSWIYTQRNGLMDIGENEYRLILKKNLLMPLAAI----LALILTPIAPEYSSTAYAV 181

Query: 178 IPL 180
            + L
Sbjct: 182 LIL 184
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 981

A DNA sequence (GBSx1040) was identified in *S.agalactiae* <SEQ ID 2999> which encodes the amino acid sequence <SEQ ID 3000>. This protein is predicted to be guanylate kinase (gmk). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13441 GB:Z99112 similar to guanylate kinase [Bacillus subtilis]
 Identities = 121/202 (59%), Positives = 155/202 (75%)

Query: 1    MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKFDYSVSMTTRPQRPGEVDGVDYFFRTRE 60
            M ERGLLIV SGPSGVGKGTVRQ IFS  D KF+YS+S+TTR  R GEV+GVDYFF+TR+
Sbjct: 41   MKERGLLIVLSGPSGVGKGTVRQAIFSQEDTKFEYSISVTTRSPREGEVNGVDYFFKTRD 100

Query: 61   EFEALIKEGQMLEYAEYVGNYYGTPLSYVNETLDKGIDVFLEIEVQGALQVKSKVPDGVF 120
            EFE +I + ++LE+AEYVGNYYGTP+ YV +TL  G DVFLEIEVQGALQV++  P+G+F
Sbjct: 101 EFEQMIADNKLLEWAEYVGNYYGTPVDYVEQTLQDGKDVFLEIEVQGALQVRNAFPEGLF 160

Query: 121 IFLTPPDLEELEERLVGRGTDSPEVIAQRIERAKEEIALMREYDYAVVNDQVSLAAERVK 180
            IFL PP L EL+ R+V RGT++  +I  R++ AK EI +M  YDY V ND V  A +++K
Sbjct: 161 IFLAPPSLSELKNRIVTRGTETDALIENRMKAAKAEIEMMDAYDYVVENDNVETACDKIK 220

Query: 181 RVIEAEHYRVDRVIGRYTNMVK 202
             ++ AEH + +RV  RY  M++
Sbjct: 221 AIVLAEHLKRERVAPRYKKMLE 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3001> which encodes the amino acid sequence <SEQ ID 3002>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13441 GB:Z99112 similar to guanylate kinase [Bacillus subtilis]
 Identities = 123/203 (60%), Positives = 157/203 (76%)

Query: 1   MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKFEYSVSMTTRPQRPGEVDGVDYFFRTRE 60
           M ERGLLIV SGPSGVGKGTVRQ IFS  D KFEYS+S+TTR  R GEV+GVDYFF+TR+
Sbjct: 41  MKERGLLIVLSGPSGVGKGTVRQAIFSQEDTKFEYSISVTTRSPREGEVNGVDYFFKTRD 100

Query: 61  EFEELIKTGQMLEYAEYVGNYYGTPLTYVNETLDKGIDVFLEIEVQGALQVKSKVPDGVF 120
           EFE++I   ++LE+AEYVGNYYGTP+ YV +TL  G DVFLEIEVQGALQV++  P+G+F
Sbjct: 101 EFEQMIADNKLLEWAEYVGNYYGTPVDYVEQTLQDGKDVFLEIEVQGALQVRNAFPEGLF 160

Query: 121 VFLTPPDLDELEDRLVGRGTDSQEVIAQRIERAKEEIALMREYDYAVVNDEVALAAERVK 180
           +FL PP L EL++R+V RGT++  +I  R++ AK EI +M  YDY V ND V  A +++K
Sbjct: 161 IFLAPPSLSELKNRIVTRGTETDALIENRMKAAKAEIEMMDAYDYVVENDNVETACDKIK 220

Query: 181 RIIETEHFRVERVIGRYDKMIKI 203
            I+  EH + ERV  RY KM+++
Sbjct: 221 AIVLAEHLKRERVAPRYKKMLEV 243
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 186/204 (91%), Positives = 197/204 (96%)

Query: 1   MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKFDYSVSMTTRPQRPGEVDGVDYFFRTRE 60
           MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKF+YSVSMTTRPQRPGEVDGVDYFFRTRE
Sbjct: 1   MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKFEYSVSMTTRPQRPGEVDGVDYFFRTRE 60

Query: 61  EFEALIKEGQMLEYAEYVGNYYGTPLSYVNETLDKGIDVFLEIEVQGALQVKSKVPDGVF 120
           EFE LIK GQMLEYAEYVGNYYGTPL+YVNETLDKGIDVFLEIEVQGALQVKSKVPDGVF
Sbjct: 61  EFEELIKTGQMLEYAEYVGNYYGTPLTYVNETLDKGIDVFLEIEVQGALQVKSKVPDGVF 120

Query: 121 IFLTPPDLEELEERLVGRGTDSPEVIAQRIERAKEEIALMREYDYAVVNDQVSLAAERVK 180
           +FLTPPDL+ELE+RLVGRGTDS EVIAQRIERAKEEIALMREYDYAVVND+V+LAAERVK
Sbjct: 121 VFLTPPDLDELEDRLVGRGTDSQEVIAQRIERAKEEIALMREYDYAVVNDEVALAAERVK 180

Query: 181 RVIEAEHYRVDRVIGRYTNMVKET 204
           R+IE EH+RV+RVIGRY  M+K T
Sbjct: 181 RIIETEHFRVERVIGRYDKMIKIT 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 982

A DNA sequence (GBSx1041) was identified in *S.agalactiae* <SEQ ID 3003> which encodes the amino acid sequence <SEQ ID 3004>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1763(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3005> which encodes the amino acid sequence <SEQ ID 3006>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1551(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 95/105 (90%), Positives = 100/105 (94%), Gaps = 1/105 (0%)

Query: 1   MMLKPSIDTLLDKVPSKYSLVILQAKRAHELEAGEKATQDFKSVKSTLRALEEIESGNVV 60
           MMLKPSIDTLLDKVPSKYSLVILQAKRAHELEAG   TQ+FKSVKSTL+ALEEIESGNVV
Sbjct: 1   MMLKPSIDTLLDKVPSKYSLVILQAKRAHELEAGATPTQEFKSVKSTLQALEEIESGNVV 60

Query: 61  IHPDPSAKRASVRARIEAERLAKEEEERKIKEQIAKEK-EDGEKI 104
           IHPDPSAKR +VRA+IEAERLAKEEEERKIKEQIAKEK E+GEKI
Sbjct: 61  IHPDPSAKREAVRAKIEAERLAKEEEERKIKEQIAKEKEEEGEKI 105
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 983

A DNA sequence (GBSx1043) was identified in *S.agalactiae* <SEQ ID 3007> which encodes the amino acid sequence <SEQ ID 3008>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3413(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13444 GB:Z99112 primosomal replication factor Y (primosomal
           protein N') [Bacillus subtilis]
 Identities = 377/807 (46%), Positives = 529/807 (64%), Gaps = 21/807 (2%)

Query: 6   AQVIVDIPLMQTDKPFSYAIPKDLEDLVQVGVRVHVPFGRGNRLLQGFVVGFRDDDELET 65
           A+VIVD+     D+PF Y IP  L+ +++ G+RV VPFG   R +QGFV   ++  +L
Sbjct: 4   AEVIVDVSTKNIDRPFDYKIPDHLKGMIKTGMRVIVPFGP--RKIQGFVTAVKEASDLSG 61

Query: 66  KDIAEV---LDFEPVLNQEQLDLADQMRHTVFSYKISILKSMLPSLLNSQYDKLLL---A 119
           K + EV   LD  PVL +E + L+  +     S+KI+ L++MLP+ L ++Y+K L
Sbjct: 62  KSVKEVEDLLDLTPVLTEELMILSSWLSDKTLSFKITALQAMLPAALKAKYEKELKIAHG 121

Query: 120 TDTLPSEDREDLFGHKTEIVFSSLSSQDAKKA-GRLIQKGFIEVQYLAKDKKTIKTEKIY 178
            D  P  +R  LF     +++S +   +  K   R +QKG I+V Y    K   K  +
Sbjct: 122 ADLPPQVER--LFSETKTLLYSDIPDHETLKLIQRHVQKGDIDVTYKVAQKTNKKMVRHI 179

Query: 179 KINRTLLEKSQ----IAARAKKRLELKEFLLENPQPGRLTALN----KQFSSPVVNFFRE 230
           + N +  E ++    ++ +A K+  +  FL+  P+  ++ A         SS  +    +
Sbjct: 180 QANASKEELAKQAEGLSRQAAKQQAILHFLISEPEGVKIPAAELCKKTDTSSATIKTLIQ 239

Query: 231 EGIIEVIEKEASRSDNYFKGILKTDFLDLNQEQAKVVKIVVDQIGKEQNKPFLLEGITGS 290
           +G+++   +E  R     K   KT+ L L  EQ    + + + +  +++K FLL G+TGS
Sbjct: 240 KGLLKESYEEVYRDPYQDKMFKKTEPLPLTDEQRAAFEPIRETLDSDEHKVFLLHGVTGS 299

Query: 291 GKTEVYLHIIDNVLKLGKTAIVLVPEISLTPQMTNRFISRFGKQVAIMHSGLSEGEKFDE 350
           GKTE+YL  I+ VL  GK AIVLVPEISLTPQM NRF  RFG QVA+MHSGLS GEK+DE
Sbjct: 300 GKTEIYLQSIEKVLAKGKEAIVLVPEISLTPQMVNRFKGRFGSQVAVMHSGLSTGEKYDE 359

Query: 351 WRKIKSGQAKVVVGARSAIFAPLENIGAIIIDEEHESTYKQESNPRYHARDVALLRAEYY 410
           WRKI   + ++VVGARSAIFAP EN+G IIIDEEHES+YKQE  PRYHA++VA+ RAE++
Sbjct: 360 WRKIHRKEVRLVVGARSAIFAPFENLGMIIIDEEHESSYKQEEMPRYHAKEVAIKRAEHH 419
```

```
Query: 411 KAVLLMGSATPSIESRARASRDVYKFLELKHRANPKARIPQVEIIDFRNFIGQQEVSNFT 470
              +++GSATP++ES ARA + VY+ L LKHR N +  +P+V ++D R  +     S F+
Sbjct: 420 SCPVVLGSATPTLESYARAQKGVYELLSLKHRVNHRV-MPEVSLVDMREELRNGNRSMFS 478

Query: 471 SYLLDKIRDRLDKKEQVVLMLNRRGYSSFIMCRDCGYVDQCPNCDISLTLHMATKTMNCH 530
             L++K+ + + K EQ VL LN+RGYSSF+MCRDCGYV QCP+CDIS+T H   + + CH
Sbjct: 479 VELMEKLEETIAKGEQAVLFLNKRGYSSFVMCRDCGYVPQCPHCDISMTYHRYGQRLKCH 538

Query: 531 YCGFEKPIPRTCPNCNSKSISYYGTGTQKAYEELLKVIPDAKILRMDVDTTRQKGGHESI 590
           YCG E+P+P TCP C S+ I ++GTGTQ+  EEL KV+P A+++RMDVDTT +KG HE +
Sbjct: 539 YCGHEEPVPHTCPECASEHIRFFGTGTQRVEEELTKVLPSARVIRMDVDTTSRKGAHEKL 598

Query: 591 LKRFGNHEADILLGTQMIAKGLDFPNVTLVGVLNADTSLNLPDFRSSERTFQLLTQVAGR 650
           L  FG  +ADILLGTQMIAKGLDFPNVTLVGVL+ADT+L++PDFRS+E+TFQLLTQV+GR
Sbjct: 599 LSAFGEGKADILLGTQMIAKGLDFPNVTLVGVLSADTTLHIPDFRSAEKTFQLLTQVSGR 658

Query: 651 AGRAEKEGEVVIQTYNPNHYAIQLAQKQDFEAFYQYEMNIRRQLGYPPYYFTVGLTLSHK 710
           AGR EK G V+IQTY P+HY+IQL +  D+E FYQ+EM  RR+  YPPYY+   +T+SH+
Sbjct: 659 AGRHEKPGHVIIQTYTPSHYSIQLTKTHDYETFYQHEMAHRREQSYPPYYYLALVTVSHE 718

Query: 711 DEEWLIRKSYEVLSLLKQGFSDKVKLLGPTPKPIARTHNLYHYQIIIKYRFEDNLELVLN 770
           +       + ++   LK       K+LGP+   PIAR  + Y YQ +IKY+ E  L  +L
Sbjct: 719 EVAKAAVTAEKIAHFLKANCGADTKILGPSASPIARIKDRYRYQCVIKYKQETQLSALLK 778

Query: 771 RLLD-MTQDKENRDLRLAIDHEPQNMM 796
           ++L+   ++ E + + ++ID  P  MM
Sbjct: 779 KILEHYKREIEQKHVMISIDMNPYMMM 805
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3009> which encodes the amino acid sequence <SEQ ID 3010>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1396(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 556/793 (70%), Positives = 659/793 (82%), Gaps = 1/793 (0%)

Query: 4   KLAQVIVDIPLMQTDKPFSYAIPKDLEDLVQVGVRVHVPFGRGNRLLQGFVVGFRDDDEL 63
           K+A VIVDIPLMQTDKPFSY IPK+L  LVQ+G RVHVPFG+GNRLLQGF++GF  +D
Sbjct: 12  KVAHVIVDIPLMQTDKPFSYGIPKELVSLVQLGSRVHVPFGKGNRLLQGFIIGFGQEDSS 71

Query: 64  ETKDIAEVLDFEPVLNQEQLDLADQMRHTVFSYKISILKSMLPSLLNSQYDKLLLATDTL 123
             K I  VLD EPVLNQEQL LADQ+R TVFSYKI++LK+M+P+LLNS YDK+L      L
Sbjct: 72  SLKLIQTVLDPEPVLNQEQLTLADQLRKTVFSYKITLLKAMIPNLLNSNYDKVLRPESGL 131

Query: 124 PSEDREDLFGHKTEIVFSSLSSQDAKKAGRLIQKGFIEVQYLAKDKKTIKTEKIYKINRT 183
              DR+ LF  K  +++S+L  +  K A + IQ G I V YLAKDKK +KTEK Y ++
Sbjct: 132 KKSDRDFLFEGKPSVLYSTLDREKEKIALKGIQAGHITVSYLAKDKKNLKTEKYYHVDLD 191

Query: 184 LLEKSQIAARAKKRLELKEFLLENPQPGRLTALNKQFSSPVVNFFREEGIIEVIEKEASR 243
            L    I++RAKKR  LK++LL + +  +L  L + FS  VV +F    +I + E+   R
Sbjct: 192 ALAVHPISSRAKKRQLLKDYLLTHTKEAKLATLYQAFSRDVVAYFVTNHLIRIDERPIDR 251

Query: 244 SDNYFKGILKTDFLDLNQEQAKVVKIVVDQIGKEQNKPFLLEGITGSGKTEVYLHIIDNV 303
           S++YF  I  + FL LN++QA  V  +V+QIGK  +KPFL+EGITGSGKTEVYLHII+ V
Sbjct: 252 SESYFDQIKPSSFLTLNEQQASAVTEIVEQIGKP-SKPFLIEGITGSGKTEVYLHIIEAV 310

Query: 304 LKLGKTAIVLVPEISLTPQMTNRFISRFGKQVAIMHSGLSEGEKFDEWRKIKSGQAKVVV 363
           LK  KTAIVLVPEISLTPQMT+RFISRFGKQVAIMHSGLS+GEKFDEWRKIK+GQAKVVV
Sbjct: 311 LKQDKTAIVLVPEISLTPQMTSRFISRFGKQVAIMHSGLSDGEKFDEWRKIKTGQAKVVV 370
```

```
Query: 364 GARSAIFAPLENIGAIIIDEEHESTYKQESNPRYHARDVALLRAEYYKAVLLMGSATPSI 423
           GARSAIF+PLE IGAIIIDEEHESTYKQESNPRYHAR+VALLRA++++AV++MGSATPSI
Sbjct: 371 GARSAIFSPLERIGAIIIDEEHESTYKQESNPRYHAREVALLRAKHHQAVVVMGSATPSI 430

Query: 424 ESRARASRDVYKFLELKHRANPKARIPQVEIIDFRNFIGQQEVSNFTSYLLDKIRDRLDK 483
           ESRARAS+ VY F++L  RANP A+IP+V I+DFR++IGQQ VSNFT YL+DKI++RL K
Sbjct: 431 ESRARASKGVYHFIQLTQRANPLAKIPEVTIVDFRDYIGQQAVSNFTPYLIDKIKERLVK 490

Query: 484 KEQVVLMLNRRGYSSFIMCRDCGYVDQCPNCDISLTLHMATKTMNCHYCGFEKPIPRTCP 543
           KEQVVLMLNRRGYSSF+MCRDCGYVD+CPNCDISLTLHM TKTMNCHYCGF+KPIP TCP
Sbjct: 491 KEQVVLMLNRRGYSSFVMCRDCGYVDKCPNCDISLTLHMDTKTMNCHYCGFQKPIPITCP 550

Query: 544 NCNSKSISYYGTGTQKAYEELLKVIPDAKILRMDVDTTRQKGGHESILKRFGNHEADILL 603
            C+S SI YYGTGTQKA++EL  VIP+AKILRMDVDTTR+K  H++IL  FG  EADILL
Sbjct: 551 ECHSNSIRYYGTGTQKAFDELQGVIPEAKILRMDVDTTRKKRSHKTILDSFGRQEADILL 610

Query: 604 GTQMIAKGLDFPNVTLVGVLNADTSLNLPDFRSSERTFQLLTQVAGRAGRAEKEGEVVIQ 663
           GTQMIAKGLDFPNVTLVGVLNADTSLNLPDFR+SE+TFQLLTQVAGRAGRA K GEV+IQ
Sbjct: 611 GTQMIAKGLDFPNVTLVGVLNADTSLNLPDFRASEKTFQLLTQVAGRAGRAHKPGEVLIQ 670

Query: 664 TYNPNHYAIQLAQKQDFEAFYQYEMNIRRQLGYPPYYFTVGLTLSHKDEEWLIRKSYEVL 723
           TYNP+HYAIQLA+KQDFEAFY+YEM+IR Q+ YPPYYFTVG+TLSH+ E  +++K+Y+V
Sbjct: 671 TYNPDHYAIQLAKKQDFEAFYRYEMSIRHQMAYPPYYFTVGITLSHRLEASVVKKAYQVT 730

Query: 724 SLLKQGFSDKVKLLGPTPKPIARTHNLYHYQIIIKYRFEDNLELVLNRLLDMTQDKENRD 783
            LLK   SD +K+LGPTPKPIARTHNLYHYQI++KYRFEDNLE  LNR+LD +Q+ +NR
Sbjct: 731 ELLKSHLSDNIKILGPTPKPIARTHNLYHYQILLKYRFEDNLEETLNRILDWSQEADNRH 790

Query: 784 LRLAIDHEPQNMM 796
           L+L ID EPQ  +
Sbjct: 791 LKLIIDCEPQQFL 803
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 984

A DNA sequence (GBSx1044) was identified in *S.agalactiae* <SEQ ID 3011> which encodes the amino acid sequence <SEQ ID 3012>. This protein is predicted to be methionyl-tRNA formyltransferase (fmt). Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1329(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13446 GB:Z99112 methionyl-tRNA formyltransferase [Bacillus subtilis]
 Identities = 155/314 (49%), Positives = 221/314 (70%), Gaps = 7/314 (2%)

Query: 1   MTKLLFMGTPDFSATVLKGILADGKYDVLAVVTQPDRAVGRKKEIKMTPVKEVALENNIP 60
            MT+++FMGTPDFS  VL+ ++ DG Y+V+ VVTQPDR  GRKK +   PVKE AL + IP
Sbjct: 1   MTRIVFMGTPDFSVPVLRTLIEDG-YEVVGVVTQPDRPKGRKKVLTPPPVKEEALRHGIP 59

Query: 61  VYQPEKLSGSPELEQLMTLGADGIVTAAFGQFLPTKLLESVGFA-INVHASLLPKYRGGA 119
            V QPEK+  + E+E+++ L  D IVTAAFGQ LP +LL+S  +  INVHASLLP+ RGGA
Sbjct: 60  VLQPEKVRLTEEIEKVLALKPDLIVTAAFGQILPKELLDSPKYGCINVHASLLPELRGGA 119

Query: 120 PIHYAIINGEKEAGVTIMEMVAKMDAGDMVSKASVEITDEDNVGTMFDRLAVVGRDLLLD 179
            PIHY+I+ G+K+ G+TIM MV K+DAGDM+SK  V+I + DNVGT+ D+L+V G  LL +
```

```
Sbjct: 120 PIHYSILQGKKKTGITIMYMVEKLDAGDMISKVEVDIEETDNVGTLHDKLSVAGAKLLSE 179

Query: 180 TLPGYLSGDIKPIPQNEEEVSFSPNISPDEERIDWNKSSRDIFNHVRGMYPWPVAHTLLE 239
           T+P  ++G I P  Q+EE+ +++PNI  ++E +DW+++  +++N +RG+ PWPVA+T L
Sbjct: 180 TVPNVIAGSISPEKQDEEKATYAPNIKREQELLDWSRTGEELYNQIRGLNPWPVAYTTLN 239

Query: 240 GNRFKLY--EVTMSEGKGSPGQVIAKTKNSLTVATG-DGAIELKSVQPAGKPRMDIKDFL 296
           G   K++  +   +     PG V+A  K  + VATG + A+ L  +QPAGK RM  +DF+
Sbjct: 240 GQNLKIWASKKIAAPTTAEPGTVVAVEKEGIIVATGNETALLLTELQPAGKKRMKGEDFV 299

Query: 297 NGVGRNLEIGDKFG 310
            G   ++E GD  G
Sbjct: 300 RGA--HVEAGDVLG 311
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3013> which encodes the amino acid sequence <SEQ ID 3014>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0730(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 217/310 (70%), Positives = 266/310 (85%)

Query: 1   MTKLLFMGTPDFSATVLKGILADGKYDVLAVVTQPDRAVGRKKEIKMTPVKEVALENNIP 60
           M KLLFMGTP FSATVLKG+L +  Y++L VVTQPDRAVGRKK+IK+TPVK++ALE+ I
Sbjct: 1   MIKLLFMGTPQFSATVLKGLLDNPAYEILGVVTQPDRAVGRKKDIKVTPVKQLALEHGIS 60

Query: 61  VYQPEKLSGSPELEQLMTLGADGIVTAAFGQFLPTKLLESVGFAINVHASLLPKYRGGAP 120
           +YQPEKLSGS EL ++M LGADGI+TAAFGQFLPT LL+SV FAINVHASLLPKYRGGAP
Sbjct: 61  IYQPEKLSGSQELIEIMGLGADGIITAAFGQFLPTILLDSVSFAINVHASLLPKYRGGAP 120

Query: 121 IHYAIINGEKEAGVTIMEMVAKMDAGDMVSKASVEITDEDNVGTMFDRLAVVGRDLLLDT 180
           IHYAI+NG+KEAGVTIMEM+ +MDAGDMV+KAS  I + DNVGT+F++LA++GRDLLLD+
Sbjct: 121 IHYAIMNGDKEAGVTIMEMIKEMDAGDMVAKASTPILETDNVGTLFEKLAIIGRDLLLDS 180

Query: 181 LPGYLSGDIKPIPQNEEEVSFSPNISPDEERIDWNKSSRDIFNHVRGMYPWPVAHTLLEG 240
           LP YLSG++KPIPQ+  + +FSPNISP+ E++DW  S++++FNH+RGM PWPVAHT LEG
Sbjct: 181 LPAYLSGELKPIPQDHSQATFSPNISPEHEKLDWTMSNQEVFNHIRGMNPWPVAHTFLEG 240

Query: 241 NRFKLYEVTMSEGKGSPGQVIAKTKNSLTVATGDGAIELKSVQPAGKPRMDIKDFLNGVG 300
            R K+YE  ++EG+G PGQV+ KTK SL +ATG GA+ L  VQPAGKP+M I DFLNG+G
Sbjct: 241 QRLKIYEAQLAEGEGLPGQVVVKTKKSLVIATGQGALSLIVVQPAGKPKMSIIDFLNGIG 300

Query: 301 RNLEIGDKFG 310
           R LE+GD  G
Sbjct: 301 RKLEVGDIIG 310
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 985

A DNA sequence (GBSx1045) was identified in *S.agalactiae* <SEQ ID 3015> which encodes the amino acid sequence <SEQ ID 3016>. This protein is predicted to be sunL protein (sun). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1677(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 222/434 (51%), Positives = 305/434 (70%), Gaps = 15/434 (3%)

Query: 7   KSARGLALMTLEEVFDKGAYSNIALNKSLKKSRLSDKDRALVTEIVYGTVARKITLEWYL 66
            K+AR  AL  L ++F   AY+NI+L+++L+ S LS  D+  VT +VYG V++K  LEWY+
Sbjct: 3   KNARQTALDVLNDIFGNDAYANISLDRNLRDSELSTVDKGFVTALVYGVVSKKALLEWYI 62

Query: 67  SHFIVDRDKLELWVYHLLLLSLYQLLYLDNIPDHAIVNDAVTIAKNRGNKKGAEKLINAV 126
            +  +    K   W   LLLL++YQ+L++D +P  A V++AV IAK R + +     INAV
Sbjct: 63  TPLLKKEPKP--WAKMLLLLTIYQVLFMDKVPISAAVDEAVKIAK-RHDGQATANFINAV 119

Query: 127 LRR-VSSETLPEIASIKRQNKRYSVAYSMPVWLVKKLIDQYGETRALAIMESLFERNKAS 185
            LR  + SE   E      + K +   YSMP  L+ K++ Q+G  R   I+ESL + +  S
Sbjct: 120 LRNFMRSEHRNE------EPKDWETKYSMPKLLLDKMVRQFGGKRTGEILESLEKPSHVS 173

Query: 186 LRVTDLSQKQTIKETLNVRDSHIAETALVADSGNFASTSFFQDGLITIQDESSQLVAPTL 245
            LR  D +      E    R S + ETAL+ADSGNF+ T  FQ G ITIQDE+SQLVAP L
Sbjct: 174 LRKIDPTV-----EIAGTRPSLLTETALIADSGNFSITEEFQTGRITIQDETSQLVAPQL 228

Query: 246 KVSGNDQVLDACSAPGGKTSHIASYLTTGAVTALDLYDHKLELVMENAKRLGLSDKIKTK 305
            ++ G ++VLDAC+APGGK++H+A YLTTG +TALDLY+HKL+L+ +NA+R  ++DKI T+
Sbjct: 229 ELEGTEEVLDACAAPGGKSTHMAQYLTTGHITALDLYEHKLDLINQNAQRQHVADKITTQ 288

Query: 306 KLDASKAHEYFLEDTFDKILVDAPCSGIGLIRRKPDIKYNKANQDFEALQEIQLSILSSV 365
            K DA+  +E F  + FD+ILVDAPCSGIGLIRRKPDI+Y K + DF  LQ+IQL IL+S
Sbjct: 289 KADATMIYENFGPEKFDRILVDAPCSGIGLIRRKPDIRYRKESSDFIDLQKIQLEILNSA 348

Query: 366 CQTLRKGGIITYSTCTIFEEENFQVIEKFLENHPNFEQVELSHTQEDIVKRGCISISPEQ 425
             ++L+K GI+ YSTCTIF+EENF V+ +FLENHPNFEQVE+S+ + +++K GC+ I+PE
Sbjct: 349 SKSLKKSGIMVYSTCTIFDEENFDVVHEFLENHPNFEQVEISNEKPEVIKEGCLFITPEM 408

Query: 426 YHTDGFFIGQVKRI 439
            YHTDGFFI + K+I
Sbjct: 409 YHTDGFFIAKFKKI 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3017> which encodes the amino acid sequence <SEQ ID 3018>. Analysis of this protein sequence reveals the following:

```
       Possible site: 42
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 208/433 (48%), Positives = 287/433 (66%), Gaps = 13/433 (3%)

Query: 7   KSTRGKALLVIEAIFDQGAYTNIALNQQLSNKALSAKDRALLTEIVYGTVSRKISLEWYL 66
            K+ R  AL V+  IF   AY NI+L++ L +  LS  D+  +T +VYG VS+K  LEWY+
Sbjct: 3   KNARQTALDVLNDIFGNDAYANISLDRNLRDSELSTVDKGFVTALVYGVVSKKALLEWYI 62

Query: 67  AHYVKDRDKLDKWVYYLLMLSLYQLTYLDKLPAHAIVNDAVGIAKNRGNKKGAEKFVNAI 126
               +K   K   W   LL+L++YQ+ ++DK+P  A V++AV IAK R + +    F+NA+
Sbjct: 63  TPLLKKEPK--PWAKMLLLLTIYQVLFMDKVPISAAVDEAVKIAK-RHDGQATANFINAV 119
```

```
Query: 127 LRQFTSHPLPDMETIKRRNKYYSVKYSLPVWLVKKLEDQFGSDRSVAIMESLFVRSKASI 186
           LR F        E      K +  KYS+P  L+ K+  QFG  R+  I+ESL   S  S+
Sbjct: 120 LRNFMRS-----EHRNEEPKDWETKYSMPKLLLDKMVRQFGGKRTGEILESLEKPSHVSL 174

Query: 187 RVTDPLKLEEVAEALDAERSLLSATGLTKASGHFAASDYFTNGDITIQDESSQLVAPTLN 246
           R  DP       E      SLL+ T L   SG+F+ ++ F  G ITIQDE+SQLVAP L
Sbjct: 175 RKIDP-----TVEIAGTRPSLLTETALIADSGNFSITEEFQTGRITIQDETSQLVAPQLE 229

Query: 247 IDGDDIILDACSAPGGKTSHIASYLKTGKVIALDLYDHKLELVKENANRLGVADNIETRK 306
           ++G + +LDAC+APGGK++H+A YL TG + ALDLY+HKL+L+ +NA R  VAD I T+K
Sbjct: 230 LEGTEEVLDACAAPGGKSTHMAQYLTTGHITALDLYEHKLDLINQNAQRQHVADKITTQK 289

Query: 307 LDAREVHRHFEKDSFDKILVDAPCSGIGLIRRKPDIKYNKESQGFNALQAIQLEILSSVC 366
            DA  ++ +F  + FD+ILVDAPCSGIGLIRRKPDI+Y KES  F  LQ IQLEIL+S
Sbjct: 290 ADATMIYENFGPEKFDRILVDAPCSGIGLIRRKPDIRYRKESSDFIDLQKIQLEILNSAS 349

Query: 367 QTLRKGGIITYSTCTIFDEENRQVIEAFLQSHPNFEQVKLNHTQADIVKDGYLIITPEQY 426
           ++L+K GI+ YSTCTIFDEEN  V+  FL++HPNFEQV++++ + +++K+G L ITPE Y
Sbjct: 350 KSLKKSGIMVYSTCTIFDEENFDVVHEFLENHPNFEQVEISNEKPEVIKEGCLFITPEMY 409

Query: 427 QTDGFFIGQVRRV 439
            TDGFFI + +++
Sbjct: 410 HTDGFFIAKFKKI 422
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 305/440 (69%), Positives = 370/440 (83%)

Query: 1   MANDWKKSARGLALMTLEEVFDKGAYSNIALNKSLKKSRLSDKDRALVTEIVYGTVARKI 60
           +A++WKKS RG AL+ +E +FD+GAY+NIALN+ L    LS KDRAL+TEIVYGTV+RKI
Sbjct: 1   LADNWKKSTRGKALLVIEAIFDQGAYTNIALNQQLSNKALSAKDRALLTEIVYGTVSRKI 60

Query: 61  TLEWYLSHFIVDRDKLELWVYHLLLLSLYQLLYLDNIPDHAIVNDAVTIAKNRGNKKGAE 120
           +LEWYL+H++ DRDKL+ WVY+LL+LSLYQL YLD +P HAIVNDAV IAKNRGNKKGAE
Sbjct: 61  SLEWYLAHYVKDRDKLDKWVYYLLMLSLYQLTYLDKLPAHAIVNDAVGIAKNRGNKKGAE 120

Query: 121 KLINAVLRRVSSETLPEIASIKRQNKRYSVAYSMPVWLVKKLIDQYGETRALAIMESLFE 180
           K +NA+LR+ +S  LP++ +IKR+NK YSV YS+PVWLVKKL DQ+G  R++AIMESLF
Sbjct: 121 KFVNAILRQFTSHPLPDMETIKRRNKYYSVKYSLPVWLVKKLEDQFGSDRSVAIMESLFV 180

Query: 181 RNKASLRVTDLSQKQTIKETLNVRDSHIAETALVADSGNFASTSFFQDGLITIQDESSQL 240
           R+KAS+RVTD  + + + E L+   S ++ T L   SG+FA++ +F +G ITIQDESSQL
Sbjct: 181 RSKASIRVTDPLKLEEVAEALDAERSLLSATGLTKASGHFAASDYFTNGDITIQDESSQL 240

Query: 241 VAPTLKVSGNDQVLDACSAPGGKTSHIASYLTTGAVTALDLYDHKLELVMENAKRLGLSD 300
           VAPTL + G+D +LDACSAPGGKTSHIASYL TG V ALDLYDHKLELV ENA RLG++D
Sbjct: 241 VAPTLNIDGDDIILDACSAPGGKTSHIASYLKTGKVIALDLYDHKLELVKENANRLGVAD 300

Query: 301 KIKTKKLDASKAHEYFLEDTFDKILVDAPCSGIGLIRRKPDIKYNKANQDFEALQEIQLS 360
            I+T+KLDA + H +F +D+FDKILVDAPCSGIGLIRRKPDIKYNK +Q F ALQ IQL
Sbjct: 301 NIETRKLDAREVHRHFEKDSFDKILVDAPCSGIGLIRRKPDIKYNKESQGFNALQAIQLE 360

Query: 361 ILSSVCQTLRKGGIITYSTCTIFEEENFQVIEKFLENHPNFEQVELSHTQEDIVKRGCIS 420
           ILSSVCQTLRKGGIITYSTCTIF+EEN QVIE FL++HPNFEQV+L+HTQ DIVK G +
Sbjct: 361 ILSSVCQTLRKGGIITYSTCTIFDEENRQVIEAFLQSHPNFEQVKLNHTQADIVKDGYLI 420

Query: 421 ISPEQYHTDGFFIGQVKRIL 440
           I+PEQY TDGFFIGQV+R+L
Sbjct: 421 ITPEQYQTDGFFIGQVRRVL 440
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 986

A DNA sequence (GBSx1046) was identified in *S.agalactiae* <SEQ ID 3019> which encodes the amino acid sequence <SEQ ID 3020>. This protein is predicted to be pppL protein. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5796(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10712 GB:AJ132604 pppL protein [Lactococcus lactis]
 Identities = 131/245 (53%), Positives = 177/245 (71%), Gaps = 4/245 (1%)

Query: 1   MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
           ME S+L+DIG +RS NQD++  + N+AG  L +LADGMGGH+AGN+AS++TV DLG  W+
Sbjct: 1   MEYSILSDIGSKRSTNQDYVGTYVNRAGYQLFLLADGMGGHKAGNVASKLTVEDLGKLWS 60

Query: 61  ETDF---SELSEIRDWMLVSIETENRKIYELGQSDDYKGMGTTIEAVAIVGDNIIFAHVG 117
           ET F   +  + +  W+   +  EN  I  LG+ D+Y+GMGTT+EA+ I G+ I+ AHVG
Sbjct: 61  ETFFDAGTPEATLEIWLRNQVRNENENIASLGKLDEYQGMGTTLEALVIKGNTIVSAHVG 120

Query: 118 DSRIGIVRQGEYHLLTSDHSLVNELVKAGQLTEEEAASHPQKNIITQSIGQANPVEPDLG 177
           DSR  ++R GE + +T+DHSLV ELV AGQ+TEEEA  HP KNIIT+S+GQ N V+ D+
Sbjct: 121 DSRTYLMRDGELNKITTDHSLVQELVDAGQITEEEAEVHPNKNIITRSLGQTNEVQADIQ 180

Query: 178 VHLLEEGDYLVVNSDGLTNMLSNADIATVLTQEK-TLDDKNQDLITLANHRGGLDNITVA 236
              L+ GD +++NSDGLTNM+S  +I  VL +E  TLD+K++ LI LAN  GGLDNITV
Sbjct: 181 ALELQAGDIILMNSDGLTNMVSTTEIMEVLEREDLTLDNKSEALIRLANEHGGLDNITVV 240

Query: 237 LVYVE 241
           L+  E
Sbjct: 241 LIKFE 245
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3021> which encodes the amino acid sequence <SEQ ID 3022>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5301(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 180/245 (73%), Positives = 220/245 (89%)

Query: 1   MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
           M+ISL TDIGQ+RSNNQDFIN+F+NK G+ L+ILADGMGGHRAGNIASEMTVTDLG +W
Sbjct: 1   MKISLKTDIGQKRSNNQDFINKFDNKKGITLVILADGMGGHRAGNIASEMTVTDLGREWV 60

Query: 61  ETDFSELSEIRDWMLVSIETENRKIYELGQSDDYKGMGTTIEAVAIVGDNIIFAHVGDSR 120
           +TDF+ELS+IRDW+  +I++EN++IY+LGQS+D+KGMGTT+EAVA+V  + I+AH+GDSR
Sbjct: 61  KTDFTELSQIRDWLFETIQSENQRIYDLGQSEDFKGMGTTVEAVALVESSAIYAHIGDSR 120

Query: 121 IGIVRQGEYHLLTSDHSLVNELVKAGQLTEEEAASHPQKNIITQSIGQANPVEPDLGVHL 180
           IG+V  G Y LLTSDHSLVNELVKAGQ+TEEEAASHPQ+NIITQSIGQA+PVEPDLGV +
Sbjct: 121 IGLVHDGHYTLLTSDHSLVNELVKAGQITEEEAASHPQRNIITQSIGQASPVEPDLGVRV 180
```

```
Query: 181 LEEGDYLVVNSDGLTNMLSNADIATVLTQEKTLDDKNQDLITLANHRGGLDNITVALVYV 240
           LE GDYLV+NSDGLTNM+SN +I T+L  + +LD+KNQ++I LAN RGGLDNIT+ALV+
Sbjct: 181 LEPGDYLVINSDGLTNMISNDEIVTILGSKVSLDEKNQEMIDLANLRGGLDNITIALVHN 240

Query: 241 ESEAV 245
           ESE V
Sbjct: 241 ESEDV 245
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 987

A DNA sequence (GBSx1047) was identified in *S.agalactiae* <SEQ ID 3023> which encodes the amino acid sequence <SEQ ID 3024>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.03   Transmembrane  346 - 362 ( 340 - 372)

----- Final Results -----
               bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9539> which encodes amino acid sequence <SEQ ID 9540> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10713 GB:AJ132604 hypothetical protein [Lactococcus lactis]
 Identities = 219/380 (57%), Positives = 284/380 (74%), Gaps = 8/380 (2%)

Query: 1   MIQIGKLFAGRYRILKSIGRGGMADVYLARDLILDNEEVAIKVLRTNYQTDQIAVARFQR 60
           MIQIGK+FA RYRI+K IGRGGMA+VY   D  L + +VAIKVLR+N++ D IA+ARFQR
Sbjct: 1   MIQIGKIFADRYRIIKEIGRGGMANVYQGEDTFLGDRKVAIKVLRSNFENDDIAIARFQR 60

Query: 61  EARAMAELTHPNIVAIRDIGEEDGQQFLVMEYVDGFDLKKYIQDNAPLSNNEVVRIMNEV 120
           EA AMAEL+HPNIV I D+GE + QQ++VME+VDG  LK+YI  NAPL+N+E + I+ E+
Sbjct: 61  EAFAMAELSHPNIVGISDVGEFESQQYIVMEFVDGMTLKQYINQNAPLANDEAIEIITEI 120

Query: 121 LSAMSLAHQKGIVHRDLKPQNILLTKKGTVKVTDFGIAVAFAETSLTQTNSMLGSVHYLS 180
           LSAM +AH  GI+HRDLKPQN+L++  GTVKVTDFGIA A +ETSLTQTN+M GSVHYLS
Sbjct: 121 LSAMDMAHSHGIIHRDLKPQNVLVSSSGTVKVTDFGIAKALSETSLTQTNTMFGSVHYLS 180

Query: 181 PEQARGSKATVQSDIYAMGIMLFEMLTGHIPYDGDSAVTIALQHFQKPLPSILAENKSVP 240
           PEQARGS ATVQSDIYA+GI+LFE+LTG IP+DGDSAV IAL+HFQ+ +PSI+  N  VP
Sbjct: 181 PEQARGSNATVQSDIYAIGIILFELLTGQIPFDGDSAVAIALKHFQENIPSIINLNPEVP 240

Query: 241 QALENIVIKATAKKLTDRYKTTYEMGRDLSTALSSTRHREPKLVFN-DTESTKTLPKVTS 299
           QALEN+VIKATAK + +RY    EM  D++T+ S  R  E KLVFN D + TK +P   +
Sbjct: 241 QALENVVIKATAKDINNRYADVEEMMTDVATSTSLDRRGEEKLVFNKDHDETKIMP--AN 298

Query: 300 TVSSLTTEQLLRNQKQAKTTEKITPDSASNDKTKSKKKASHRLLGTIMKLFFALCVVGII 359
            ++   T+ L+   K+    EK   +S++ +  K+K K S +  G I+ L   L V+G
Sbjct: 299 LINPYDTKPLI--DKKTDDQEKAQSESSTTENNKNKNKKSKK--GLIISLVVLLLVIGGG 354

Query: 360 VFAYKILVSPTTIRVPDVSN 379
            FA+ +  +PT ++VP+V+N
Sbjct: 355 AFAWAV-STPTNVKVPNVTN 373
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3025> which encodes the amino acid sequence <SEQ ID 3026>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.60   Transmembrane  349 - 365 ( 340 - 370)

----- Final Results -----
               bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA10713 GB:AJ132604 hypothetical protein [Lactococcus lactis]
 Identities = 209/378 (55%), Positives = 273/378 (71%), Gaps = 8/378 (2%)

Query: 1   MIQIGKLFAGRYRILKSIGRGGMADVYLANDLILDNEDVAIKVLRTNYQTDQVAVARFQR 60
           MIQIGK+FA RYRI+K IGRGGMA+VY   D  L +  VAIKVLR+N++ D +A+ARFQR
Sbjct: 1   MIQIGKIFADRYRIIKEIGRGGMANVYQGEDTFLGDRKVAIKVLRSNFENDDIAIARFQR 60

Query: 61  EARAMAELNHPNIVAIRDIGEEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEV 120
           EA AMAEL+HPNIV I D+GE + QQ++VME+VDG  LK+YI  +APL+N+E + I+ E+
Sbjct: 61  EAFAMAELSHPNIVGISDVGEFESQQYIVMEFVDGMTLKQYINQNAPLANDEAIEIITEI 120

Query: 121 LSAMTLAHQKGIVHRDLKPQNILLTKEGVVKVTDFGIAVAFAETSLTQTNSMLGSVHYLS 180
           LSAM +AH  GI+HRDLKPQN+L++  G VKVTDFGIA A +ETSLTQTN+M GSVHYLS
Sbjct: 121 LSAMDMAHSHGIIHRDLKPQNVLVSSSGTVKVTDFGIAKALSETSLTQTNTMFGSVHYLS 180

Query: 181 PEQARGSKATIQSDIYAMGIMLFEMLTGHIPYDGDSAVTIALQHFQKPLPSIIEENHNVP 240
           PEQARGS AT+QSDIYA+GI+LFE+LTG IP+DGDSAV IAL+HFQ+ +PSII  N  VP
Sbjct: 181 PEQARGSNATVQSDIYAIGIILFELLTGQIPFDGDSAVAIALKHFQENIPSIINLNPEVP 240

Query: 241 QALENVVIRATAKKLSDRYGSTFEMSRDLMTALSYNRSRERKIIF-ENVESTKPLPKVAS 299
           QALENVVI+ATAK +++RY    EM  D+ T+ S +R  E K++F ++ + TK +P
Sbjct: 241 QALENVVIKATAKDINNRYADVEEMMTDVATSTSLDRRGEEKLVFNKDHDETKIMPANLI 300

Query: 300 GPTASVKLSPPTPTVLTQESRLDQTNQTDALQPPTKKKKSGRFLGTLFKILFSFFIVGVA 359
            P  +  L         QE    +++ T+  +   KK K G  +  +  +L    ++G
Sbjct: 301 NPYDTKPLIDKKTD--DQEKAQSESSTTENNKNKNKKSKKGLIISLVVLLL----VIGGG 354

Query: 360 LFTYLILTKPTSVKVPNV 377
            F + + T PT+VKVPNV
Sbjct: 355 AFAWAVST-PTNVKVPNV 371
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 390/643 (60%), Positives = 480/643 (73%), Gaps = 29/643 (4%)

Query: 1   MIQIGKLFAGRYRILKSIGRGGMADVYLARDLILDNEEVAIKVLRTNYQTDQIAVARFQR 60
           MIQIGKLFAGRYRILKSIGRGGMADVYLA DLILDNE+VAIKVLRTNYQTDQ+AVARFQR
Sbjct: 1   MIQIGKLFAGRYRILKSIGRGGMADVYLANDLILDNEDVAIKVLRTNYQTDQVAVARFQR 60

Query: 61  EARAMAELTHPNIVAIRDIGEEDGQQFLVMEYVDGFDLKKYIQDNAPLSNNEVVRIMNEV 120
           EARAMAEL HPNIVAIRDIGEEDGQQFLVMEYVDG DLK+YIQ++APLSNNEVVRIM EV
Sbjct: 61  EARAMAELNHPNIVAIRDIGEEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEV 120

Query: 121 LSAMSLAHQKGIVHRDLKPQNILLTKKGTVKVTDFGIAVAFAETSLTQTNSMLGSVHYLS 180
           LSAM+LAHQKGIVHRDLKPQNILLTK+G VKVTDFGIAVAFAETSLTQTNSMLGSVHYLS
Sbjct: 121 LSAMTLAHQKGIVHRDLKPQNILLTKEGVVKVTDFGIAVAFAETSLTQTNSMLGSVHYLS 180

Query: 181 PEQARGSKATVQSDIYAMGIMLFEMLTGHIPYDGDSAVTIALQHFQKPLPSILAENKSVP 240
           PEQARGSKAT+QSDIYAMGIMLFEMLTGHIPYDGDSAVTIALQHFQKPLPSI+ EN +VP
Sbjct: 181 PEQARGSKATIQSDIYAMGIMLFEMLTGHIPYDGDSAVTIALQHFQKPLPSIIEENHNVP 240

Query: 241 QALENIVIKATAKKLTDRYKTTYEMGRDLSTALSSTRHREPKLVFNDTESTKTLPKVTS- 299
           QALEN+VI+ATAKKL+DRY +T+EM RDL TALS  R RE K++F + ESTK LPKV S
Sbjct: 241 QALENVVIRATAKKLSDRYGSTFEMSRDLMTALSYNRSRERKIIFENVESTKPLPKVASG 300

Query: 300 ----------TVSSLTTEQLLRNQKQAKTTEKITPDSASNDKTKSKKKASHRLLGTIMKL 349
                     T + LT E  L    Q   T+ + P +        KKK S R LGT+ K+
```

```
Sbjct: 301 PTASVKLSPPTPTVLTQESRL---DQTNQTDALQPPT--------KKKKSGRFLGTLFKI 349

Query: 350 FFALCVVGIIVFAYKILVSPTTIRVPDVSNKTVAQAKMTLENSGLKVGAIRNIESDSVSE 409
            F+  +VG+ +F Y IL  PT+++VP+V+  ++  AK  L + GLKVG IR IESD+V+E
Sbjct: 350 LFSFFIVGVALFTYLILTKPTSVKVPNVAGTSLKVAKQELYDVGLKVGKIRQIESDTVAE 409

Query: 410 GLVVKTDPAAGRSRREGAKVNLYIATPNKSFTLGNYKEHNYKDILKDL-QGKGVKKSLIK 468
           G VV+TDP AG ++R+G+ + LY++  NK F + NYK  +Y++ +  L +  GV KS IK
Sbjct: 410 GNVVRTDPKAGTAKRQGSSITLYVSIGNKGFDMENYKGLDYQEAMNSLIETYGVPKSKIK 469

Query: 469 VKRKINNDYTTGTILAQSLPEGTSFNPDGNKKLTLTVAVNDPMIMPDVTGMTVGEVIETL 528
           ++R + N+Y   T+++QS   G  FNP+G  K+TL+VAV+D + MP VT  +  + + TL
Sbjct: 470 IERIVTNEYPENTVISQSPSAGDKFNPNGKSKITLSVAVSDTITMPMVTEYSYADAVNTL 529

Query: 529 TDLGLDADNLVFYQMQNGV---YQTVVTPPSSSKIASQDPYYGGEVGLRRGDKVKLYLLG 585
           T LG+DA  +  Y   +     +  + +P S + ++ Q PYYG  + L    ++ LYL
Sbjct: 530 TALGIDASRIKAYVPSSSSATGFVPIHSPSSKAIVSGQSPYYGTSLSLSDKGEISLYLYP 589

Query: 586 SKTTNNSSSTPIDSSASSSTGTTTSDSVSSSTDASTSDSSSTS 628
            +T ++SSS+   SS SSS  ++ +DS +  ++   S S +TS
Sbjct: 590 EETHSSSSSS---SSTSSSNSSSINDSTAPGSNTELSPSETTS 629
```

SEQ ID 3024 (GBS297) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 6; MW 75kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 4; MW 100.2kDa) and in Figure 159 (lane 2-4; MW 100kDa). GBS297-GST was purified as shown in Figure 223, lane 3. GBS297-His was purified as shown in Figure 203, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 988

A DNA sequence (GBSx1048) was identified in *S.agalactiae* <SEQ ID 3027> which encodes the amino acid sequence <SEQ ID 3028>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.91    Transmembrane    60 -  76 (  50 -  90)
   INTEGRAL    Likelihood = -7.43    Transmembrane     7 -  23 (   3 -  25)
   INTEGRAL    Likelihood = -5.68    Transmembrane    27 -  43 (  24 -  46)

----- Final Results -----
              bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03323 GB:AB035448 hypothetical protein [Staphylococcus
           aureus]
 Identities = 53/230 (23%), Positives = 104/230 (45%), Gaps = 14/230 (6%)

Query: 5   QFFLLVEAVVLVMGLMKILSDDWTSFIFILAL--ILLALRF-YNNDSRHNFLLTTSLLLL 61
           Q  ++  A++++     I  +    F+ +L L  +L+ + + Y +  R         LL+
Sbjct: 9   QMLIIFTALMIIANFYYIFFEK-IGFLLVLLLGCVLVYVGYLYFHKIRGLLAFWIGALLI 67

Query: 62  FLIFMLNPY-IIAAVVFAVLYVLINHFSQVKKKNRYALIQFKNHQLDVKTTRNQWLGTDQ 120
               + N Y II   VF +L ++     + K K   A  +       +K    +W G  +
Sbjct: 68  AFTLLSNKYTIIILFVFLLLLIVRYLIHKFKPKKVVATDEVMTSPSFIK---QKWFGEQR 124

Query: 121 HESDFYAFEDINIIRISGTDTIDLTNVIVSGQDNVIIIQKVFGDTKVLVPLDVAVKADIS 180
                Y +ED+ I    G   IDLT      ++N I+++ + G  +V++P++  +   ++
```

```
Sbjct: 125 TPVYVYKWEDVQIQHGIGDLHIDLTKAANIKENNTIVVRHILGKVQVILPVNYNINLHVA 184

Query: 181 SVYGSVQYFDFEEYDLRNESIKLSQ--EEEYYLLKRVKLVVNTIAGKVEV 228
           + YGS  Y + + Y + N +I + +  + + Y    V + V+T  G VEV
Sbjct: 185 AFYGST-YVNEKSYKVENNNIHIEEMMKPDNY---TVNIYVSTFIGDVEV 230
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3029> which encodes the amino acid sequence <SEQ ID 3030>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.92   Transmembrane    44 -  60 (  36 -  64)
   INTEGRAL    Likelihood = -8.76   Transmembrane    69 -  85 (  66 - 105)
   INTEGRAL    Likelihood = -8.70   Transmembrane    24 -  40 (  20 -  42)
   INTEGRAL    Likelihood = -6.64   Transmembrane    88 - 104 (  85 - 105)

----- Final Results -----
              bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03323 GB:AB035448 hypothetical protein [Staphylococcus
           aureus]
 Identities = 41/187 (21%), Positives = 85/187 (44%), Gaps = 22/187 (11%)

Query: 47   FILILVL--ILLALRF-YNQDSRNNFLLTVSLLFLFLIFMLNPYIIMAVLLGIVYIFINH 103
             F+L+L+L  +L+ + + Y    R      +  L +    + N Y I+ + + ++ + +
Sbjct: 33   FLLVLLLGCVLVYVGYLYFHKIRGLLAFWIGALLIAFTLLSNKYTIIILFVFLLLLIV-- 90

Query: 104 FSQVKKKNRFALIRFKEEKIEVNNT--------KHQWIGTANYESDYYCFDDINIIRISG 155
                    R+ + +FK +K+   +         K +W G        Y ++D+ I    G
Sbjct: 91   --------RYLIHKFKPKKVVATDEVMTSPSFIKQKWFGEQRTPVYVYKWEDVQIQHGIG 142

Query: 156 NDTVDLTNVIVTGMDNIIVIRKIFGNTTILVPIDVTVTLDVSSIYGSVDFFRCQQYDLRN 215
           +  +DLT       +N IV+R I G   +++P++  + L V++ YGS  +   + Y + N
Sbjct: 143 DLHIDLTKAANIKENNTIVVRHILGKVQVILPVNYNINLHVAAFYGST-YVNEKSYKVEN 201

Query: 216 ESIKFKE 222
            +I  +E
Sbjct: 202 NNIHIEE 208
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 137/211 (64%), Positives = 175/211 (82%)

Query: 1    MKKFQFFLLVEAVVLVMGLMKILSDDWTSFIFILALILLALRFYNNDSRHNFLLTTSLLL 60
            MKKFQFFLL+E ++L MG+M IL +D +SFI IL LILLALRFYN DSR+NFLLT SLL
Sbjct: 18   MKKFQFFLLIECILLAMGIMTILDNDLSSFILILVLILLALRFYNQDSRNNFLLTVSLLF 77

Query: 61   LFLIFMLNPYIIAAVVFAVLYVLINHFSQVKKKNRYALIQFKNHQLDVKTTRNQWLGTDQ 120
            LFLIFMLNPYII AV+  ++Y+ INHFSQVKKKNR+ALI+FK  +++V  T++QW+GT
Sbjct: 78   LFLIFMLNPYIIMAVLLGIVYIFINHFSQVKKKNRFALIRFKEEKIEVNNTKHQWIGTAN 137

Query: 121 HESDFYAFEDINIIRISGTDTIDLTNVIVSGQDNVIIIQKVFGDTKVLVPLDVAVKADIS 180
           +ESD+Y F+DINIIRISG DT+DLTNVIV+G DN+I+I+K+FG+T +LVP+DV V  D+S
Sbjct: 138 YESDYYCFDDINIIRISGNDTVDLTNVIVTGMDNIIVIRKIFGNTTILVPIDVTVTLDVS 197

Query: 181 SVYGSVQYFDFEEYDLRNESIKLSQEEEYYL 211
           S+YGSV +F  ++YDLRNESIK  + +   L
Sbjct: 198 SIYGSVDFFRCQQYDLRNESIKFKETDNQSL 228
```

SEQ ID 3028 (GBS66) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 4; MW 25kDa) and in Figure 7 (lane 2; MW 24.7kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 989

A DNA sequence (GBSx1049) was identified in *S.agalactiae* <SEQ ID 3031> which encodes the amino acid sequence <SEQ ID 3032>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.41   Transmembrane   47 -  63 (  40 -  72)
   INTEGRAL    Likelihood = -9.98   Transmembrane    9 -  25 (   5 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54570 GB:AJ006393 histidine kinase [Streptococcus pneumoniae]
 Identities = 159/334 (47%), Positives = 239/334 (70%), Gaps = 5/334 (1%)

Query: 1   MKKHHYFLAFFYGSVIIFAICFVIIDSLGVNL-VHLYQTSRLWLIEQLIFSIFFLSLAVT 59
           MKK  Y +      + +F     +++ L  +  + L+   +    E+ +F +   S+++T
Sbjct: 1   MKKQAYVIIALTSFLFVFFFSHSLLEILDFDWSIFLHDVEKT---EKFVFLLLVFSMSMT 57

Query: 60  ILLLLTWFLLDDNSKRQINHNLRRILNNQSINVTDDGTEISTNIQRLSKKMNLMTASLQS 119
            LL L W  +++ S R++  NL+R+L  Q +    D  ++  + + LS K+NL+T +LQ
Sbjct: 58  CLLALFWRGIEELSLRKMQANLKRLLAGQEVVQVAD-PDLDASFKSLSGKLNLLTEALQK 116

Query: 120 KENSRILKSQEIVKQERKRIARDLHDTVSQDLFAASMVLSGIAQNVSQLDVDQVGSQLLA 179
            EN  + + +EI+++ERKRIARDLHDTVSQ+LFAA M+LSGI+Q   +LD +++ +QL +
Sbjct: 117 AENQSLAQEEEIIEKERKRIARDLHDTVSQELFAAHMILSGISQQALKLDREKMQTQLQS 176

Query: 180 VEEMLQHAQNDLRILLLHLRPVELENKTLSEGFRMILKELTDKSDIEVVYHESILTLPKK 239
           V  +L+ AQ DLR+LLLHLRPVELE K+L EG +++LKEL DKSD+ V   +++  LPKK
Sbjct: 177 VTAILETAQKDLRVLLLHLRPVELEQKSLIEGIQILLKELEDKSDLRVSLKQNMTKLPKK 236

Query: 240 IEDNIFRIGQEFISNTLKHSQASRLEVYLNQTENELQLKMIDNGIGFDMDSVYDLSYGLK 299
           IE++IFRI QE ISNTL+H+QAS L+VYL QT+ ELQLK++DNGIGF + S+ DLSYGL+
Sbjct: 237 IEEHIFRILQELISNTLRHAQASCLDVYLYQTDVELQLKVVDNGIGFQLGSLDDLSYGLR 296

Query: 300 NIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQ 333
           NI++RVED+AG +QLL+ P +G+A+DIR+PL+++
Sbjct: 297 NIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDK 330
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2991> which encodes the amino acid sequence <SEQ ID 2992>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.22   Transmembrane   49 -  65 (  42 -  70)
   INTEGRAL    Likelihood = -6.58   Transmembrane    8 -  24 (   5 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 218/337 (64%), Positives = 276/337 (81%), Gaps = 3/337 (0%)
```

```
Query: 1   MKKHHYFLAFFYGSVIIFAICFVIIDSLGVNLVHLYQTSRLWLIEQLIFSIFFLSLAVTI 60
           MKK +Y L + Y ++ I +I FV++D+LG+   +L   + LW +E+L FSI  L ++VT+
Sbjct: 1   MKKRYYALVWLYSTITILSIVFVVMDNLGITFNYL--RNHLWQVERLGFSILLLIVSVTL 58

Query: 61  LLLLTWFLLDDNSKRQINHNLRRILNNQSINVTDDGTEISTNIQRLSKKMNLMTASLQSK 120
           LLLL W ++DDNSKR IN NL+ ILNN+ + + D+ +EI+TN+ RLSKKM+ +TA++Q K
Sbjct: 59  LLLLLWIIMDDNSKRNINQNLKYILNNRRLYL-DETSEINTNLSRLSKKMSHLTANMQKK 117

Query: 121 ENSRILKSQEIVKQERKRIARDLHDTVSQDLFAASMVLSGIAQNVSQLDVDQVGSQLLAV 180
           E++ IL SQE+VKQERKRIARDLHDTVSQ+LFA+S++LSGI+ ++ QLD  Q+ +QL  V
Sbjct: 118 ESAYILDSQEVVKQERKRIARDLHDTVSQELFASSLILSGISMSLEQLDKTQLQTQLTTV 177

Query: 181 EEMLQHAQNDLRILLLHLRPVELENKTLSEGFRMILKELTDKSDIEVVYHESILTLPKKI 240
           E MLQ+AQNDLRILLLHLRP EL N+TLSEG  MILKELTDKSDIEV+Y E+I  LPK +
Sbjct: 178 EAMLQNAQNDLRILLLHLRPTELANRTLSEGLHMILKELTDKSDIEVIYKETIAQLPKTM 237

Query: 241 EDNIFRIGQEFISNTLKHSQASRLEVYLNQTENELQLKMIDNGIGFDMDSVYDLSYGLKN 300
           EDN+FRI QEFISNTLKH++ASR+EVYLNQT  ELQLKMID+G+GFDMD V DLSYGLKN
Sbjct: 238 EDNLFRIAQEFISNTLKHAKASRIEVYLNQTSTELQLKMIDDGVGFDMDQVRDLSYGLKN 297

Query: 301 IEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQSEDK 337
           IEDRV DLAGNL L+SQ GKGV+MDIRLP+V   +D+
Sbjct: 298 IEDRVNDLAGNLHLISQKGKGVSMDIRLPIVKGDDDE 334
```

A related GBS gene <SEQ ID 8701> and protein <SEQ ID 8702> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      14.69
GvH: Signal Score (-7.5): -4.31
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -11.41 threshold:  0.0
   INTEGRAL    Likelihood =-11.41   Transmembrane   47 -  63 (  40 -  72)
   INTEGRAL    Likelihood = -9.98   Transmembrane    9 -  25 (   5 -  36)
   PERIPHERAL  Likelihood =  3.61     146
 modified ALOM score:   2.78

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
52.5/77.6% over 288aa
                                                    Streptococcus pneumoniae
  GP|5830526| histidine kinase Insert characterized

ORF00320(433 - 1302 of 1617)
GP|5830526|emb|CAB54570.1||AJ006393(43  -  331  of  331)  histidine  kinase  {Streptococcus
pneumoniae}
%Match = 28.6
%Identity = 52.4  %Similarity = 77.6
Matches = 152  Mismatches = 64  Conservative Sub.s = 73

252       282       312       342       372       402       432       462
QEEEYTF*NVSN*L*TLSLES*G*S*MKKHHYFLAFFYGSVIIFAICFVIIDSLGVNLVHLYQTSRLWLIEQLIFSIFFL
                                       :  :|   |       :::        :   |:::| :: :
                            MKKQAYVIIALTSFLFVFFFSHSLLEILDFDWSIFLHDVEKTEKFVFLLLVF
                                    10        20        30        40        50

492       522       552       582       612       642       672       702
SLAVTILLLLLTWFLLDDNSKRQINHNLRRILNNQSINVTDDGTEISTNIQRLSKKMNLMTASLQSKENSRILKSQEIVKQ
```

```
|:::| || | |  ::: | |::  ||:|:|  | :    |  ::  : : || |:||:| :||  ||  : : :||:::
SMSMTCLLALFWRGIEELSLRKMQANLKRLLAGQEVVQVAD-PDLDASFKSLSGKLNLLTEALQKAENQSLAQEEEIIEK
                70        80        90       100       110       120       130

732       762       792       822       852       882       912       942
ERKRIARDLHDTVSQDLFAASMVLSGIAQNVSQLDVDQVGSQLLAVEEMLQHAQNDLRILLLHLRPVELENKTLSEGFRM
|||||||||||||||:|||| |:||||:|   :|| ::: :|| :|  :|: || |||:||||||||||| |:| || ::
ERKRIARDLHDTVSQELFAAHMILSGISQQALKLDREKMQTQLQSVTAILETAQKDLRVLLLHLRPVELEQKSLIEGIQI
                150       160       170       180       190       200       210

972      1002      1032      1062      1092      1122      1152      1182
ILKELTDKSDIEVVYHESILTLPKKIEDNIFRIGQEFISNTLKHSQASRLEVYLNQTENELQLKMIDNGIGFDMDSVYDL
:|||| ||||: |   :::  ||||||::|||| ||:|||||:|:||| |:||| ||: |||||::|||||| : |: ||
LLKELEDKSDLRVSLKQNMTKLPKKIEEHIFRILQELISNTLRHAQASCLDVYLYQTDVELQLKVVDNGIGFQLGSLDDL
                230       240       250       260       270       280       290

1212      1242      1272      1302      1332      1362      1392      1422
SYGLKNIXDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQSEDKNG*NKNCTC**P*DGSSRFKKFFKLTS*C*SNR*GLK
||||:|| :||||:|| :|||: | :|:|:|||:||:::
SYGLRNIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDKE
                310       320       330
```

SEQ ID 8702 (GBS31) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 8; MW 64kDa). It was also expressed as GBS31d in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 8-10; MW 59kDa) and in Figure 187 (lane 8; MW 59kDa). GBS31d was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 11-13; MW 34kDa) and in Figure 182 (lane 11; MW 34kDa). Purified GBS31d-GST is shown in lane 3 of Figure 237.

Example 990

A DNA sequence (GBSx1050) was identified in *S.agalactiae* <SEQ ID 3033> which encodes the amino acid sequence <SEQ ID 3034>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54571 GB:AJ006393 response regulator [Streptococcus pneumoniae]
 Identities = 154/209 (73%), Positives = 184/209 (87%)

Query: 8   IKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMPEMD 67
           +KI+LVDDHEMVRLGLKS+ +LQ DVEV+GEASNG +GI  ALELRPDV+VMD+VMPEM+
Sbjct: 1   MKILLVDDHEMVRLGLKSYFDLQDDVEVVGEASNGSQGIDLALELRPDVIVMDIVMPEMN 60

Query: 68  GVEATLALLKDWPEAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKVSRG 127
           G++ATLA+LK+WPEA IL++TSYLDNEKI PV++AGAKGYMLKTSSA E+L+A+ KV+ G
Sbjct: 61  GIDATLAILKEWPEAKILIVTSYLDNEKIMPVLDAGAKGYMLKTSSADELLHAVSKVAAG 120

Query: 128 EQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVKTHV 187
           E AIE EV KK++ H     LHE LTARERD+L L+AKGY+NQRIAD+LFISLKTVKTHV
Sbjct: 121 ELAIEQEVSKKVEYHRNHMELHEELTARERDVLQLIAKGYENQRIADDLFISLKTVKTHV 180

Query: 188 SNILGKLNVADRTQAVVYAFQHHLVPQDD 216
           SNIL KL V+DRTQA VYAFQHHLV Q++
Sbjct: 181 SNILAKLEVSDRTQAAVYAFQHHLVGQEE 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2995> which encodes the amino acid sequence <SEQ ID 2996>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3094(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 175/212 (82%), Positives = 192/212 (90%)

Query: 5   MDKIKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMP 64
            M KIK++LVDDHEMVR+GLKSFLNLQAD++V+GEASNG EG+  AL L+PDV+VMDLVMP
Sbjct: 3   MSKIKVILVDDHEMVRMGLKSFLNLQADIDVVGEASNGREGVDLALALKPDVLVMDLVMP 62

Query: 65  EMDGVEATLALLKDWPEAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKV 124
            E+ GVEATL +LK W EA +LVLTSYLDNEKIYPVI+AGAKGYMLKTSSAAEILNAIRKV
Sbjct: 63  ELGGVEATLEVLKKWKEAKVLVLTSYLDNEKIYPVIDAGAKGYMLKTSSAAEILNAIRKV 122

Query: 125 SRGEQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVK 184
            S+GE AIE EVDKKIKAHD+ P LHE LTARE DIL+LLAKGYDNQ IADELFISLKTVK
Sbjct: 123 SKGELAIETEVDKKIKAHDQHPDLHEELTAREYDILHLLAKGYDNQTIADELFISLKTVK 182

Query: 185 THVSNILGKLNVADRTQAVVYAFQHHLVPQDD 216
            THVSNIL KL V DRTQAVVYAF+HHLVPQDD
Sbjct: 183 THVSNILAKLEVGDRTQAVVYAFRHHLVPQDD 214
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 991

A DNA sequence (GBSx1051) was identified in *S.agalactiae* <SEQ ID 3035> which encodes the amino acid sequence <SEQ ID 3036>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1688(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB08166 GB:Z94864 putative peptidyl-prolyl cis-trans isomerase
          [Schizosaccharomyces pombe]
 Identities = 81/174 (46%), Positives = 109/174 (62%), Gaps = 30/174 (17%)

Query: 288 IKTNHGDMTVKLFPDHAPKTVANFIGLAKQGYYDGIIFHRIIPDFMIQGGDPTGTGMGGE 347
            ++T+ G + ++L+ +HAPKT  NF  LAK+GYYDG+IFHR+IPDF+IQGGDPTGTG GG
Sbjct: 6   LQTSLGKILIELYTEHAPKTCQNFYTLAKEGYYDGVIFHRVIPDFVIQGGDPTGTGRGGT 65

Query: 348 SIYGESFEDEFSEELYNV-RGALSMANAGPNTNGSQFFIVQNTKIPYAKKELERGGWPTP 406
            SIYG+ F+DE   +L++   G LSMANAGPNTN SQFFI   T  P
Sbjct: 66  SIYGDKFDDEIHSDLHHTGAGILSMANAGPNTNSSQFFI---TLAP-------------- 108

Query: 407 IAELYAGQGGTPHLDRRHSVFGQLVDQSSFEVLDEIAAVETGSQDKPLEDVVIL 460
                      TP LD +H++FG++V  S   V   +  + T S D+P+E + I+
Sbjct: 109 ----------TPWLDGKHTIFGRVV--SGLSVCKRMGLIRTDSSDRPIEPLKII 150
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3037> which encodes the amino acid sequence <SEQ ID 3038>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2175(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 381/464 (82%), Positives = 422/464 (90%)

Query: 1   MDAKTKYKAKKIKAVFFDIDDTLRVKDTGYMPPSILKVFKALKDKGIVVGIASGRARYGV 60
           MDAK KYKAKKIK VFFDIDDTLRVKDTGYMP SI +VFKALK KGI+VGIASGRARYGV
Sbjct: 5   MDAKLKYKAKKIKMVFFDIDDTLRVKDTGYMPESIQRVFKALKAKGILVGIASGRARYGV 64

Query: 61  PKEVQDLNADYCVKLNGAYVKDKDKNIIFHRPIPAEYVEQYKKWADTVGIKYGLAGRHEA 120
           P+EVQDL+ADYCVKLNGAYVKD  K IIF  PIPA+ V  YKKWAD +GI YG+AGRHEA
Sbjct: 65  PQEVQDLHADYCVKLNGAYVKDDAKTIIFQAPIPADVVVAYKKWADDMGIFYGMAGRHEA 124

Query: 121 VLSDRDDLVNDAIDIVYSDLEVNPDFNKEHDIYQMWTFEDKGDSLHLPEPLAEHLRLIRW 180
           VLS R+D++++AID VY+ LEV PD+N+ HD+YQMWTFEDKGD L LP  LAEHLRL+RW
Sbjct: 125 VLSARNDMISNAIDNVYAQLEVCPDYNEYHDVYQMWTFEDKGDGLQLPAELAEHLRLVRW 184

Query: 181 HDHSSDVVLKGTSKALGVSKVVEHLGLKPENILVFGDELNDLELFDYAGLAVAMGVSHPE 240
           HD+SSDVVLKGTSKALGVSKVV+HLGLKPENILVFGDELNDLELFDYAG+++AMGVSHP
Sbjct: 185 HDNSSDVVLKGTSKALGVSKVVDHLGLKPENILVFGDELNDLELFDYAGISIAMGVSHPL 244

Query: 241 AQKKADFITKKVEEDGILYALEELGLIEKELTFPQVDIENTEGPVAVIKTNHGDMTVKLF 300
            Q+KADFITKKVEEDGILYALEELGLI+KEL FPQ+D+ N +GP A IKTNHGDMT+ LF
Sbjct: 245 LQEKADFITKKVEEDGILYALEELGLIDKELQFPQLDLPNHKGPKATIKTNHGDMTLVLF 304

Query: 301 PDHAPKTVANFIGLAKQGYYDGIIFHRIIPDFMIQGGDPTGTGMGGESIYGESFEDEFSE 360
           PDHAPKTVANF+GLAK+GYYDGIIFHRIIP+FMIQGGDPTGTGM G+SIYGESFEDEFS+
Sbjct: 305 PDHAPKTVANFLGLAKEGYYDGIIFHRIIPEFMIQGGDPTGTGMCGQSIYGESFEDEFSD 364

Query: 361 ELYNVRGALSMANAGPNTNGSQFFIVQNTKIPYAKKELERGGWPTPIAELYAGQGGTPHL 420
           ELYN+RGALSMANAGPNTNGSQFFIVQN+KIPYAKKELERGGWP PIA  YA  GGTPHL
Sbjct: 365 ELYNLRGALSMANAGPNTNGSQFFIVQNSKIPYAKKELERGGWPAPIAASYAANGGTPHL 424

Query: 421 DRRHSVFGQLVDQSSFEVLDEIAAVETGSQDKPLEDVVILTIEV 464
           DRRH+VFGQLVD++SF+VLD IA VETG+QDKP EDV+I TIEV
Sbjct: 425 DRRHTVFGQLVDETSFQVLDLIAGVETGAQDKPKEDVIIETIEV 468
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 992

A DNA sequence (GBSx1052) was identified in *S.agalactiae* <SEQ ID 3039> which encodes the amino acid sequence <SEQ ID 3040>. This protein is predicted to be ribosomal protein S1 (rpsA). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3126(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07066 GB:AP001518 polyribonucleotide nucleotidyltransferase
           (general stress protein 13) [Bacillus halodurans]
 Identities = 46/120 (38%), Positives = 71/120 (58%), Gaps = 11/120 (9%)

Query: 8   KIGDKLKGTVTGIRPYGAFVSLEDGRTGLIHISEIKTGYIDNIYDVLSVGDEVYVQVIDV 67
           ++G  ++G VTGI+P+GAFV+++D + GL+HISE+  G++ +I DVLSVGDEV V+++ V
Sbjct: 5   EVGSIVEGKVTGIKPFGAFVAIDDQKQGLVHISEVAHGFVKDINDVLSVGDEVKVKILSV 64

Query: 68  DEFTQKASLSLRTLEEERHHIQH-----------RHRFSNNRLKIGFKPLEENLPSWVEE 116
           DE + K SLS+R  +E                  R          GF  LE+ L  W+++
Sbjct: 65  DEESGKISLSIRATQEAPERPARAPKPRPAGGGGRKPQKGQSQGQGFNTLEDKLKEWLKQ 124
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3041> which encodes the amino acid sequence <SEQ ID 3042>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1832(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 78/115 (67%), Positives = 100/115 (86%)

Query: 7   MKIGDKLKGTVTGIRPYGAFVSLEDGRTGLIHISEIKTGYIDNIYDVLSVGDEVYVQVID 66
           MKIGDKL GT+TGI+PYGAFV+LE+G TGLIHISEIKTG+ID+I  +L++G++V VQVID
Sbjct: 1   MKIGDKLHGTITGIKPYGAFVALENGTTGLIHISEIKTGFIDDIDQLLAIGNQVLVQVID 60

Query: 67  VDEFTQKASLSLRTLEEERHHIQHRHRFSNNRLKIGFKPLEENLPSWVEEGLAYL 121
           +DE+++K SLS+RTL EE+ H  HRHR+SN+R KIGF+PLEE LP W+EE L +L
Sbjct: 61  IDEYSKKPSLSMRTLAEEKQHFFHRHRYSNSRHKIGFRPLEEQLPQWIEESLQFL 115
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 993

A DNA sequence (GBSx1053) was identified in *S.agalactiae* <SEQ ID 3043> which encodes the amino acid sequence <SEQ ID 3044>. This protein is predicted to be pyruvate formate-lyase 2 activating enzyme (pflA). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2889(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76934 GB:AE000469 probable pyruvate formate lyase activating
           enzyme 2 [Escherichia coli K12]
 Identities = 90/251 (35%), Positives = 142/251 (55%), Gaps = 16/251 (6%)

Query: 8   VFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQKMVPETMR--------------- 52
           +FNIQ +S++DG GIRT VF KGCP  CPWCANPES      +T+R
Sbjct: 24  IFNIQRYSLNDGEGIRTVVFFKGCPHLCPWCANPESISGKIQTVRREAKCLHCAKCLRDA 83
```

```
Query: 53  -DAITNESVIVGEEKSVDDIIEEVLKDIDFYEESGGGITLSGGEIFAQFEFAKAILKRAK 111
            +  +     +G + S+D +  EV+KD  F+  SGGG+TLSGGE+  Q EFA   L+R +
Sbjct: 84  DECPSGAFERIGRDISLDALEREVMKDDIFFRTSGGGVTLSGGEVLMQAEFATRFLQRLR 143

Query: 112 SLGIHTAIETTAYTRHEQFIDLIQYVDFIYTDLKHYNSLKHQEKTMVKNASIIKNIHYAF 171
              G+  AIET       + + L +  D +  DLK  ++ + ++   +    +++N+
Sbjct: 144 LWGVSCAIETAGDAPASKLLPLAKLCDEVLFDLKIMDATQARDVVKMNLPRVLENLRLLV 203

Query: 172 ANGKTIVLRIPVIPNFNDSLEDAEEFACLFDRLDIRQVQLLPFHQFGQNKYQLLNRQYEM 231
           + G  ++ R+P+IP F  S E+ ++   +   L+IRQ+ LLPFHQ+G+ KY+LL + + M
Sbjct: 204 SEGVNVIPRLPLIPGFTLSRENMQQALDVLIPLNIRQIHLLPFHQYGEPKYRLLGKTWSM 263

Query: 232 EEIAALHPEDL 242
           +E+ A    D+
Sbjct: 264 KEVPAPSSADV 274
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3045> which encodes the amino acid sequence <SEQ ID 3046>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2209(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 187/255 (73%), Positives = 220/255 (85%)

Query: 4   EKGIVFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQKMVPETMRDAITNESVIVG 63
           ++GIVFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQ+  PE M  +    + IVG
Sbjct: 3   DRGIVFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQQKAPEQMLTSDGLNTKIVG 62

Query: 64  EEKSVDDIIEEVLKDIDFYEESGGGITLSGGEIFAQFEFAKAILKRAKSLGIHTAIETTA 123
           EEK+VD++IEEVLKD+DFYEESGGG+TLSGGEIFAQF+FA A+LK AK+ G+HTAIETTA
Sbjct: 63  EEKTVDEVIEEVLKDLDFYEESGGGMTLSGGEIFAQFDFALALLKAAKAAGLHTAIETTA 122

Query: 124 YTRHEQFIDLIQYVDFIYTDLKHYNSLKHQEKTMVKNASIIKNIHYAFANGKTIVLRIPV 183
           + +HEQF+ L+ YVDFIYTDLKHYN L+HQ+ T V+N  IIKNIHYAF  GK IVLRIPV
Sbjct: 123 FAKHEQFVTLVDYVDFIYTDLKHYNQLRHQKVTGVRNDLIIKNIHYAFQAGKEIVLRIPV 182

Query: 184 IPNFNDSLEDAEEFACLFDRLDIRQVQLLPFHQFGQNKYQLLNRQYEMEEIAALHPEDLL 243
           IP FNDSL+DA+ F+ LF++L+I QVQLLPFHQFG+NKY+LL R+YEM E+ A HPEDL
Sbjct: 183 IPQFNDSLDDAKAFSELFNQLEIDQVQLLPFHQFGENKYKLLGREYEMAEVKAYHPEDLA 242

Query: 244 DYQAIFSKYNIHCYF 258
           DYQA+F  +NIHCYF
Sbjct: 243 DYQAVFLNHNIHCYF 257
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 994

A DNA sequence (GBSx1054) was identified in *S.agalactiae* <SEQ ID 3047> which encodes the amino acid sequence <SEQ ID 3048>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1762(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9299> which encodes amino acid sequence <SEQ ID 9300> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC74366 GB:AE000226 putative DEOR-type transcriptional
          regulator [Escherichia coli K12]
 Identities = 74/177 (41%), Positives = 113/177 (63%), Gaps = 1/177 (0%)

Query: 2   NRLENIISLVSQYQKIDVNTLSELLQVSKVTIRKDLDKLEGKGLLHREHGYAVLNSGDDL 61
            +R + I+ +V    ++ V  L++   VS+VTIR+DL+ LE    L R HG+AV    DD+
Sbjct: 3   SRQQTILQMVIDQGQVSVTDLAKATGVSEVTIRQDLNTLEKLSYLRRAHGFAVSLDSDDV 62

Query: 62  NVRLSFNHKTKKEIAALAANMVSDNDTILIESGSTCALLAENICQTKRNVTILTNSCFIA 121
              R+  N+  K+E+A  AA++V   +TI IE+GS+ ALLA  + + K+NVTI+T S +IA
Sbjct: 63  ETRMMSNYTLKRELAEFAASLVQPGETIFIENGSSNALLARTLGEQKKNVTIITVSSYIA 122

Query: 122 NYLREYDSCQIVLLGGEYQSSSQVTVGPLLKKMISLFHVSLAFVGTDGFDPKTRIYG 178
            + L++   C+++LLGG YQ   S+  VGPL ++ I   H S AF+G DG+ P+T   G
Sbjct: 123 HLLKD-APCEVILLGGVYQKKSESMVGPLTRQCIQQVHFSKAFIGIDGWQPETGFTG 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3049> which encodes the amino acid sequence <SEQ ID 3050>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/171 (76%), Positives = 150/171 (87%)

Query: 1   MNRLENIISLVSQYQKIDVNTLSELLQVSKVTIRKDLDKLEGKGLLHREHGYAVLNSGDD 60
            MNRLE II LVSQ +KIDVN+LSE L VSKVTIRKDLDKLE KGLL REHGYAVLNSGDD
Sbjct: 2   MNRLERIIQLVSQKKKIDVNSLSEQLDVSKVTIRKDLDKLESKGLLRREHGYAVLNSGDD 61

Query: 61  LNVRLSFNHKTKKEIAALAANMVSDNDTILIESGSTCALLAENICQTKRNVTILTNSCFI 120
            LNVRLS+N+  K+ IA  AA +V DNDTI+IESGSTCALLAE +CQTKRN+ ++TNSCFI
Sbjct: 62  LNVRLSYNYNIKRRIAEKAAELVQDNDTIMIESGSTCALLAEVLCQTKRNIKVITNSCFI 121

Query: 121 ANYLREYDSCQIVLLGGEYQSSSQVTVGPLLKKMISLFHVSLAFVGTDGFD 171
            ANY+R+Y SCQI+LLGG YQ +S+VTVGPLLK+MISLFHV+  FVGTDGF+
Sbjct: 122 ANYIRQYSSCQIILLGGYYQPNSEVTVGPLLKEMISLFHVNRVFVGTDGFN 172
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 995

A DNA sequence (GBSx1055) was identified in *S.agalactiae* <SEQ ID 3051> which encodes the amino acid sequence <SEQ ID 3052>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1672(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG04879 GB:AE004578 probable transcriptional regulator
           [Pseudomonas aeruginosa]
 Identities = 20/70 (28%), Positives = 40/70 (56%)

Query: 6   GFMGRDLMRSEVAQEMANAADEVIILTDSSKFNQTALVEQLPLSTVSQVITDKHPNSEIA 65
           G M   +  +E+A+ M   A ++ ++ DSSK  + AL +  PLS +++++ D+ P  E+
Sbjct: 179 GAMDFSIEEAEIARAMIAQARQLTVIADSSKLGRRALFQVFPLSRINRLVVDRKPTGELW 238

Query: 66  NLFQEAEITI 75
              Q+A + +
Sbjct: 239 EALQQARVEV 248
```

There is also homology to SEQ ID 3050.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 996

A DNA sequence (GBSx1056) was identified in *S.agalactiae* <SEQ ID 3053> which encodes the amino acid sequence <SEQ ID 3054>. This protein is predicted to be transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0904(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9541> which encodes amino acid sequence <SEQ ID 9542> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04499 GB:AP001509 transcriptional regulator [Bacillus halodurans]
 Identities = 98/309 (31%), Positives = 178/309 (56%), Gaps = 1/309 (0%)

Query: 6   ERQKLLAKVAYLYYMEGKSQSEIANELGIYRTTISRMLAKAREEGLVRIEISDFNPEIFQ 65
           E ++L+ KVA LYY EG +Q+++A ++G+ R  IS++L KA+E+G+V I I D N    +
Sbjct: 5   EERRLIVKVASLYYFEGWTQAQVAKKIGVSRPVISKLLNKAKEQGIVEIYIKDENIHTVE 64

Query: 66  LESYFKSKYHLKDIEIVSSRKDSDTSEIEKDLAHVAAAMIRKKIKENDKVGIAWGRTLSK 125
           LE   + KYHLK+  +V +        I++ +    +  + K IK  D +GI+WG T+S
Sbjct: 65  LEQRLEKKYHLKEAIVVPT-SGLTQDMIKRAIGKATSYYVSKNIKGMDSIGISWGTTVSS 123

Query: 126 VVEAMRPHPVSQVSFVPLAGGPSHINARYHVNTLVYEMSRRFQGSCTFINATLVQENANL 185
            V+        ++  +PL GG        H N L YE++++    C+++ A  + E   L
Sbjct: 124 FVQEYPYEQHRELKVIPLVGGMGRKFVELHSNLLAYELAKKMNCECSYLYAPAMVEAKEL 183

Query: 186 AKGILTSKYFEGLMDNWEKLDVAIVGVGGKPKSNEQQWLDLLNQDDFQCLDEEAAVGEIT 245
            + ++ S+    +++    + +A+VG+G   K +  + ++ L ++D   L +  AVG+++
Sbjct: 184 KERLIQSEDIASVLEEGRNVKMAVVGIGSPFKGSTMKVMNYLKEEDIATLKKIGAVGDMS 243

Query: 246 CRFFNHSGDPVNQHLAKRTIGITLEQLQKVPNRIAVAHGNYKAAALLAVLKKGYINHLVT 305
            RF++  G P++  L +  IGI L++L+++P  I V+ G +K  ++ A LK GY++ LVT
Sbjct: 244 SRFYDALGQPIDHPLNELVIGIDLDELKRIPIVIGVSEGAHKVDSVEAALKGGYLDVLVT 303

Query: 306 DFSTALNIL 314
```

```
           D STA +++
Sbjct: 304 DDSTAQSLI 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3055> which encodes the amino acid sequence <SEQ ID 3056>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2123(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 165/324 (50%), Positives = 238/324 (72%), Gaps = 1/324 (0%)

Query: 3   MKLERQKLLAKVAYLYYMEGKSQSEIANELGIYRTTISRMLAKAREEGLVRIEISDFNPE 62
           MK ER++LLAKVAYL+Y++GKSQ+ I+ E+ IYRTT+ RMLAKA+EEG+VRIEI+D++ +
Sbjct: 1   MKEERRRLLAKVAYLHYVQGKSQTLISKEMNIYRTTVCRMLAKAKEEGIVRIEIADYDAD 60

Query: 63  IFQLESYFKSKYHLKDIEIVSSRKDSDTSEIEKDLAHVAAAMIRKKIKENDKVGIAWGRT 122
           +F LE Y + +Y L+ +++V ++ +    +   ++A  AA + R  +K+ DK+G++WG T
Sbjct: 61  LFALEEYVRQQYGLEKLDLVPNQVEDTPMDTLTNVAKTAAEVFRHVVKDGDKIGLSWGAT 120

Query: 123 LSKVVEAMRPHPVSQVSFVPLAGGPSHINARYHVNTLVYEMSRRFQGSCTFINATLVQEN 182
           LS +++ + P  +  V   PLAGGPSHINA+YHVNTLVY ++R F G+  F+NA ++QE+
Sbjct: 121 LSCLMDELNPKAMKDVFIYPLAGGPSHINAKYHVNTLVYRLARIFHGNSAFMNAMVIQED 180

Query: 183 ANLAKGILTSKYFEGLMDNWEKLDVAIVGVGGKPKSNEQ-QWLDLLNQDDFQCLDEEAAV 241
            +LAKGIL SKYF  ++ +W++LD+A+VG+GG+P S EQ QW DLL   D   L  E AV
Sbjct: 181 KHLAKGILQSKYFNDILTSWDQLDLALVGIGGEPNSLEQSQWRDLLTSSDHDQLKYEKAV 240

Query: 242 GEITCRFFNHSGDPVNQHLAKRTIGITLEQLQKVPNRIAVAHGNYKAAALLAVLKKGYIN 301
           GE+ CRFF+ +G PV   L  RTIGI+LEQL++VP  +AVA G +KA A+LA LK G+IN
Sbjct: 241 GEVCCRFFDQAGQPVYTGLQDRTIGISLEQLRRVPKTMAVATGKHKAKAILAALKAGFIN 300

Query: 302 HLVTDFSTALNILRLDKDTFVDTI 325
           +LVTD  T L +L LD+D  ++ +
Sbjct: 301 YLVTDKETMLAVLALDEDIDLNNV 324
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 997

A DNA sequence (GBSx1057) was identified in *S.agalactiae* <SEQ ID 3057> which encodes the amino acid sequence <SEQ ID 3058>. This protein is predicted to be PTS enzyme III cel (celC). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9543> which encodes amino acid sequence <SEQ ID 9544> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA23551 GB:M93570 PTS enzyme III cel [Escherichia coli]
 Identities = 42/102 (41%), Positives = 70/102 (68%)

Query: 4    EIIVADQIIMGLILNAGDAKQHIYQALKLAKEGNFAESKIEIELADSALLEAHNLQTQFL 63
            E+   ++++MGLI+N+G A+   Y ALK AK+G+FA +K  ++ +  AL EAH +QT+ +
Sbjct: 13   EVEELEEVVMGLIINSGQARSLAYAALKQAKQGDFAAAKAMMDQSRMALNEAHLVQTKLI 72

Query: 64   AQEAGGTRTDISALFIHSQDHLMTSITEINLIKEIIDLRQEL 105
              +AG  +  +S + +H+QDHLMTS+    LI E+I+L ++L
Sbjct: 73   EGDAGEGKMKVSLVLVHAQDHLMTSMLARELITELIELHEKL 114
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3059> which encodes the amino acid sequence <SEQ ID 3060>. Analysis of this protein sequence reveals the following:

```
     Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC74806 GB:AE000268 PEP-dependent phosphotransferase enzyme III
           for cellobiose, arbutin, and salicin [Escherichia coli]
 Identities = 39/97 (40%), Positives = 66/97 (67%)

Query: 7    DQIIMGLILNAGDAKQHIYQALKCAKEDDYATSEKEMALADDALLEAHNLQTQFLAQEAS 66
            ++++MGLI+N+G A+   Y ALK AK+ D+A ++  M  +  AL EAH +QT+ +  +A
Sbjct: 18   EEVVMGLIINSGQARSLAYAALKQAKQGDFAAAKAMMDQSRMALNEAHLVQTKLIEGDAG 77

Query: 67   GNKSEITALFVHSQDHLMTTITEINLIKEIIDLRKEL 103
              K +++ + VH+QDHLMT++    LI E+I+L ++L
Sbjct: 78   EGKMKVSLVLVHAQDHLMTSMLARELITELIELHEKL 114
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 81/103 (78%), Positives = 94/103 (90%)

Query: 3    MEIIVADQIIMGLILNAGDAKQHIYQALKLAKEGNFAESKIEIELADSALLEAHNLQTQF 62
            M++IV DQIIMGLILNAGDAKQHIYQALK AKE ++A S+ E+ LAD ALLEAHNLQTQF
Sbjct: 1    MQVIVPDQIIMGLILNAGDAKQHIYQALKCAKEDDYATSEKEMALADDALLEAHNLQTQF 60

Query: 63   LAQEAGGTRTDISALFIHSQDHLMTSITEINLIKEIIDLRQEL 105
            LAQEA G +++I+ALF+HSQDHLMT+ITEINLIKEIIDLR+EL
Sbjct: 61   LAQEASGNKSEITALFVHSQDHLMTTITEINLIKEIIDLRKEL 103
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 998

A DNA sequence (GBSx1058) was identified in *S.agalactiae* <SEQ ID 3061> which encodes the amino acid sequence <SEQ ID 3062>. This protein is predicted to be PTS system, cellobiose-specific IIB component (celA). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF94440 GB:AE004207 PTS system, cellobiose-specific IIB
           component [Vibrio cholerae]
 Identities = 46/100 (46%), Positives = 62/100 (62%)

Query: 1    MIKIGLFCAAGFSTGMLVNNMKIAADKEGIEAHIEAYSQGKIADYAKDLDVALLGPQVSY 60
            M KI L C+AG ST MLV  M+ AA+ +GIE  I+A S     +  ++ DV LLGPQV +
Sbjct: 1    MKKILLCCSAGMSTSMLVKKMQQAAESKGIECKIDALSVNAFEEAIQEYDVCLLGPQVRF 60

Query: 61   TLDKSKSICDEYGVPIAVIPMADYGMLDGVKVLKLALSLL 100
             L++ +   DEYG  IA I    YGM+ G +VL+ AL L+
Sbjct: 61   QLEELRKTADEYGKNIAAISPQAYGMMKGDEVLQQALDLI 100
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3063> which encodes the amino acid sequence <SEQ ID 3064>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF94440 GB:AE004207 PTS system, cellobiose-specific IIB
           component [Vibrio cholerae]
 Identities = 43/100 (43%), Positives = 58/100 (58%)

Query: 8    MIKIGLFCAAGFSTGMLVNNMKVAAEKKGIDCQIEAYAQGKLADYAPLLDVALLGPQVAY 67
            M KI L C+AG ST MLV  M+ AAE KGI+C+I+A +     +     DV LLGPQV +
Sbjct: 1    MKKILLCCSAGMSTSMLVKKMQQAAESKGIECKIDALSVNAFEEAIQEYDVCLLGPQVRF 60

Query: 68   TLDKSEAICKDNDIPIAVIPMADYGMLDGNKVLDLALSLV 107
             L++      +    IA I    YGM+ G++VL  AL L+
Sbjct: 61   QLEELRKTADEYGKNIAAISPQAYGMMKGDEVLQQALDLI 100
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 79/101 (78%), Positives = 92/101 (90%)

Query: 1    MIKIGLFCAAGFSTGMLVNNMKIAADKEGIEAHIEAYSQGKIADYAKDLDVALLGPQVSY 60
            MIKIGLFCAAGFSTGMLVNNMK+AA+K+GI+  IEAY+QGK+ADYA  LDVALLGPQV+Y
Sbjct: 8    MIKIGLFCAAGFSTGMLVNNMKVAAEKKGIDCQIEAYAQGKLADYAPLLDVALLGPQVAY 67

Query: 61   TLDKSKSICDEYGVPIAVIPMADYGMLDGVKVLKLALSLLE 101
            TLDKS++IC +  +PIAVIPMADYGMLDG KVL LALSL++
Sbjct: 68   TLDKSEAICKDNDIPIAVIPMADYGMLDGNKVLDLALSLVK 108
```

SEQ ID 3062 (GBS180) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 4; MW 12.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 2; MW 37.6kDa).

The GBS180-GST fusion product was purified (Figure 204, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 298), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 999

A DNA sequence (GBSx1059) was identified in *S.agalactiae* <SEQ ID 3065> which encodes the amino acid sequence <SEQ ID 3066>. This protein is predicted to be pts system, cellobiose-specific iic component (celB). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.68   Transmembrane  346 - 362 ( 334 - 374)
   INTEGRAL    Likelihood = -9.77   Transmembrane  182 - 198 ( 178 - 205)
   INTEGRAL    Likelihood = -8.65   Transmembrane   29 -  45 (  27 -  50)
   INTEGRAL    Likelihood = -6.53   Transmembrane  140 - 156 ( 134 - 161)
   INTEGRAL    Likelihood = -4.78   Transmembrane  292 - 308 ( 289 - 312)
   INTEGRAL    Likelihood = -4.41   Transmembrane  397 - 413 ( 395 - 416)
   INTEGRAL    Likelihood = -2.97   Transmembrane   77 -  93 (  72 -  93)
   INTEGRAL    Likelihood = -2.97   Transmembrane  228 - 244 ( 222 - 246)

----- Final Results -----
               bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA17390 GB:U07818 cellobiose phosphotransferase enzyme II''
           [Bacillus stearothermophilus]
 Identities = 160/415 (38%), Positives = 251/415 (59%), Gaps = 13/415 (3%)

Query: 15  KFVNMRGIIALKDGMLAILPLTVVGSLFLILGQLPFKGLNQAIANVFGPEWTEPFMQVYS 74
            K    R + A++DG++  +PL ++GSLFLI+G LP  G N+ +A  FG  W +  +
Sbjct: 18  KIAEQRHLQAIRDGIILSMPLLIIGSLFLIVGFLPIPGYNEWMAKWFGEHWLDKLLYPVG 77

Query: 75  GTFAIMGLISCFAIAYAYAKNSSVEPLPAGVLSLSSFFILMKSSYIPVKGEA------IA 128
             TF IM L+  F +AY  A+   V+ L AG +SL++F +L     +P   E       ++
Sbjct: 78  ATFDIMALVVSFGVAYRLAEKYKVDALSAGAISLAAF-LLATPYQVPFTPEGAKETIMVS 136

Query: 129 DAISKVWFGGQGIIGAIIIGLVVGAIYTWFIQHHIVIKMPEQVPQAIAKQFEAMIPAFVI 188
              I   W G +G+  A+I+ +V   IY   IQ +IVIK+P+ VP A+A+ F A+IP   +
Sbjct: 137 GGIPVQWVGSKGLFVAMILAIVSTEIYRKIIQKNIVIKLPDGVPPAVARSFVALIPGAAV 196

Query: 189 FLLSMIVYLIAKVTTGGTFIEMIYDIIQVPLQGLTGSLYGAIGIAFFISFLWWFGVHGQS 248
            ++  +  LI ++T   +F  ++  ++  PL  L GS++GAI     +  LW  G+HG +
Sbjct: 197 LVVVWVARLILEMTPFESFHNIVSVLLNKPLSVLGGSVFGAIVAVLLVQLLWSTGLHGAA 256

Query: 249 VVNGIVTALLLSNLDANKSLLAAN-RLTLDNGAHIVTQQFLDSFLILSGSGITFGLVIAM 307
           +V G++  + LS +D N+ +   N    L N   ++TQQF D ++ + GSG T  L + M
Sbjct: 257 IVGGVMGPIWLSLMDENRMVFQQNPNAELPN---VITQQFFDLWIYIGGSGATLALALTM 313

Query: 308 LFAAKSKQYKALGKVAAFPAIFNVNEPIVFGFPIVMNPVMFLPFILVPVLAALIVYGAIA 367
           +F A+S+Q K+LG++A  P IFN+NEPI FG PIVMNP++ +PFILVPV+  ++ Y A+A
Sbjct: 314 MFRARSRQLKSLGRLAIAPGIFNINEPITFGMPIVMNPLLIIPFILVPVVLVVVSYAAMA 373

Query: 368 VGFMQPFSGVTLPWSTPAIISGFMVGGWQ--GALVQIVILAISTAVYFPFFKIQD 420
            G +   SGV +PW+TP +ISG++  G +  G+++QIV   I+ A+Y+PFF I D
Sbjct: 374 TGLVAKPSGVAVPWTTPIVISGYLATGGKISGSILQIVNFFIAFAIYYPFFSIWD 428
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2215> which encodes the amino acid sequence <SEQ ID 2216>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.92   Transmembrane  347 - 363 ( 335 - 373)
   INTEGRAL    Likelihood = -7.59   Transmembrane   29 -  45 (  27 -  50)
   INTEGRAL    Likelihood = -7.38   Transmembrane  182 - 198 ( 179 - 204)
   INTEGRAL    Likelihood = -5.68   Transmembrane  398 - 414 ( 395 - 420)
   INTEGRAL    Likelihood = -4.99   Transmembrane  293 - 309 ( 291 - 314)
```

```
    INTEGRAL    Likelihood = -3.61   Transmembrane  140 - 156 ( 134 - 160)
    INTEGRAL    Likelihood = -2.60   Transmembrane  229 - 245 ( 229 - 246)
    INTEGRAL    Likelihood = -0.75   Transmembrane   72 -  88 (  72 -  88)

 ----- Final Results -----
                bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 366/428 (85%), Positives = 402/428 (93%), Gaps = 1/428 (0%)

Query: 1   MSKFDSQKIITPIMKFVNMRGIIALKDGMLAILPLTVVGSLFLILGQLPFKGLNQAIANV 60
           M+K + Q II PIM FVNMRGIIALKDGMLAILPLTVVGSLFLI GQ+PF+G+N AIA+V
Sbjct: 1   MAKMNMQNIIKPIMTFVNMRGIIALKDGMLAILPLTVVGSLFLIAGQIPFQGVNDAIASV 60

Query: 61  FGPEWTEPFMQVYSGTFAIMGLISCFAIAYAYAKNSSVEPLPAGVLSLSSFFILMKSSYI 120
           FG +WTEPFMQVY GTFAIMGLISCFAI Y+YAKNS VEPLP+GVLSLS+FFIL++SSY+
Sbjct: 61  FGADWTEPFMQVYHGTFAIMGLISCFAIGYSYAKNSGVEPLPSGVLSLSAFFILLRSSYV 120

Query: 121 PVKGEAIADAISKVWFGGQGIIGAIIIGLVVGAIYTWFIQHHIVIKMPEQVPQAIAKQFE 180
           P +GEAI DAISKVWFGGQGIIGAI+IGL VGA+YT FI+ HIVIKMP+QVPQAIAKQFE
Sbjct: 121 PAEGEAIGDAISKVWFGGQGIIGAIVIGLTVGAVYTTFIRRHIVIKMPDQVPQAIAKQFE 180

Query: 181 AMIPAFVIFLLSMIVYLIAK-VTTGGTFIEMIYDIIQVPLQGLTGSLYGAIGIAFFISFL 239
           AMIPAFVIF LSM+VY+IAK VT GGTFIEMIYD+IQVPLQGLTGSLYGA+GIAFFISFL
Sbjct: 181 AMIPAFVIFTLSMLVYIIAKSVTGGGTFIEMIYDVIQVPLQGLTGSLYGALGIAFFISFL 240

Query: 240 WWFGVHGQSVVNGIVTALLLSNLDANKSLLAANRLTLDNGAHIVTQQFLDSFLILSGSGI 299
           WWFGVHGQSVVNGIVTALLLSNLDANK+L+AA  L+LD GAHIVTQQFLDSFLILSGSGI
Sbjct: 241 WWFGVHGQSVVNGIVTALLLSNLDANKALMAAGELSLDKGAHIVTQQFLDSFLILSGSGI 300

Query: 300 TFGLVIAMLFAAKSKQYKALGKVAAFPAIFNVNEPIVFGFPIVMNPVMFLPFILVPVLAA 359
           TFGLV+AM+FAAKSKQYKALGKVAAFPA+FNVNEP+VFGFPIVMNPVMFLPFILVPVLAA
Sbjct: 301 TFGLVVAMIFAAKSKQYKALGKVAAFPALFNVNEPVVFGFPIVMNPVMFLPFILVPVLAA 360

Query: 360 LIVYGAIAVGFMQPFSGVTLPWSTPAIISGFMVGGWQGALVQIVILAISTAVYFPFFKIQ 419
           L VYGAIA+GFMQPF+GVTLPWSTPAIISGFMVGGWQGA+VQI+IL +ST VYFPFFKIQ
Sbjct: 361 LTVYGAIAIGFMQPFAGVTLPWSTPAIISGFMVGGWQGAIVQILILIMSTLVYFPFFKIQ 420

Query: 420 DNITYKNE 427
           DN+ Y+NE
Sbjct: 421 DNMAYQNE 428
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1000

A DNA sequence (GBSx1060) was identified in *S.agalactiae* <SEQ ID 3067> which encodes the amino acid sequence <SEQ ID 3068>. This protein is predicted to be formate acetyltransferase 2 (pflB). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5049(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC73910 GB:AE000184 putative formate acetyltransferase
```

```
               [Escherichia coli K12]
 Identities = 414/805 (51%), Positives = 555/805 (68%), Gaps = 14/805 (1%)

Query: 25  LTERMYSYRDKVLD-KKPFIDAERAILVTEAYQKHQEKPNVLKRAYMLQNILEKMTIYID 83
            L++R+ ++++ ++   KP +  ERA   TE YQ+H +KP  ++RA  L + L   TI+I
Sbjct: 9   LSDRIKAHKNALVHIVKPPVCTERAQHYTEMYQQHLDKPIPVRRALALAHHLANRTIWIK 68

Query: 84  DETMIVGNQASSDKDAPIFPEYTLEFVVNELDLFEKRDGDVFYITEETKEQIRNIAPFWE 143
             + +I+GNQAS  + APIFPEYT+ ++  E+D    R G  F ++EE K  +  + P+W
Sbjct: 69  HDELIIGNQASEVRAAPIFPEYTVSWIEKEIDDLADRPGAGFAVSEENKRVLHEVCPWWR 128

Query: 144 NNNLRARAGVMLPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLLEEGLIGFEKKARKA 203
               ++ R   M  +E +  + TG    EG M SGDAHLAVN+  LLE+GL G  ++  +
Sbjct: 129 GQTVQDRCYGMFTDEQKGLLATGIIKAEGNMTSGDAHLAVNFPLLLEKGLDGLREEVAER 188

Query: 204 KADLDLTKPESIDKYHFYDSILITIEAVKTYAERFAILAKKQAKTANAK-RRQELLDIAS 262
            ++ ++LT  E +    F  +I I + AV  + ERFA LA++ A T   + RR ELL +A
Sbjct: 189 RSRINLTVLEDLHGEQFLKAIDIVLVAVSEHIERFAALAREMAATETRESRRDELLAMAE 248

Query: 263 ICERVPYYPAETFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMYPYVKSDLEAGRETE 322
             C+ + + P +TF +A+Q  +FIQ ILQIESNGHS+S+GR DQY+YPY + D+E  +  +
Sbjct: 249 NCDLIAHQPPQTFWQALQLCYFIQLILQIESNGHSVSFGRMDQYLYPYYRRDVELNQTLD 308

Query: 323 -DSIVERLTNLWIKTITINKVRSQAHTFSSAGSPLYQNVTIGGQTR---HKEDAVNPLSF 378
             +  +E L + W+K + +NK+RS +H+ +SAGSPLYQNVTIGGQ        DAVNPLS+
Sbjct: 309 REHAIEMLHSCWLKLLEVNKIRSGSHSKASAGSPLYQNVTIGGQNLVDGQPMDAVNPLSY 368

Query: 379 LVLKSVAQTHLPQPNLTVRYHANLDKSFMNEAIEVMKLGFGMPAFNNDEIIIPSFIKKGV 438
             +L+S  +    QPNL+VRYHA +   F++  ++V++ GFGMPAFNNDEI+IP FIK G+
Sbjct: 369 AILESCGRLRSTQPNLSVRYHAGMSNDFLDACVQVIRCGFGMPAFNNDEIVIPEFIKLGI 428

Query: 439 SEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPKVLLITMNDGIDPASGKRFAP---- 494
              +DAYDY+AIGC+ETAV GKWGYRCTGMS+INF +V+L  +  G D  SGK F P
Sbjct: 429 EPQDAYDYAAIGCIETAVGGKWGYRCTGMSFINFARVMLAALEGGHDATSGKVFLPQEKA 488

Query: 495 -SYGHFTQMTSYKELKEAWDKTLRYLTRMSVIVENAIDISLEREVPDILCSALTDDCIGR 553
             S G+F    ++ E+ +AWD  +RY TR S+ +E  +D  LE  V DILCSAL DDCI R
Sbjct: 489 LSAGNFN---NFDEVMDAWDTQIRYYTRKSIEIEYVVDTMLEENVHDILCSALVDDCIER 545

Query: 554 GKHLKEGGAVYDYISGLQVGIANLSDSLAALKKLVFEEKRLTTLEVWQALQSDYAGPRGE 613
             K +K+GGA YD++SGLQVGIANL +SLAA+KKLVFE+  +   ++  AL  D+ G   E
Sbjct: 546 AKSIKQGGAKYDWVSGLQVGIANLGNSLAAVKKLVFEQGAIGQQQLAAALADDFDGLTHE 605

Query: 614 EIRQMLINEAPKYGNDDDYADSLVRECYDVYVEEIAKYPNTRYGRGPIGGIRYSGTSSIS 673
            ++RQ LIN APKYGNDDD  D+L+   Y  Y++E+ +Y N RYGRGP+GG  Y+GTSSIS
Sbjct: 606 QLRQRLINGAPKYGNDDDTVDTLLARAYQTYIDELKQYHNPRYGRGPVGGNYYAGTSSIS 665

Query: 674 ANVGQGRGTLATPDGRHAGTPLAEGCSPSHNMDKKGPTSVLKSVSKLPTDEIVGGVLLNQ 733
            ANV  G  T+ATPDGR A TPLAEG SP+   D  GPT+V+ SV KLPT  I+GGVLLNQ
Sbjct: 666 ANVPFGAQTMATPDGRKAHTPLAEGASPASGTDHLGPTAVIGSVGKLPTAAILGGVLLNQ 725

Query: 734 KVNPQTLAKEEDKQKLIALLRTFFNRLHGYHIQYNVVSRETLIDAQKHPEKHRDLIVRVA 793
            K+NP TL  E DKQKL+ LLRTFF    G+HIQYN+VSRETL+DA+KHP+++RDL+VRVA
Sbjct: 726 KLNPATLENESDKQKLMILLRTFFEVHKGWHIQYNIVSRETLLDAKKHPDQYRDLVVRVA 785

Query: 794 GYSAFFNVLSKATQDDIIARTEHAL 818
            GYSAFF  LS   QDDIIARTEH L
Sbjct: 786 GYSAFFTALSPDAQDDIIARTEHML 810
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3069> which encodes the amino acid sequence <SEQ ID 3070>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4763(Affirmative) < succ>
```

```
  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 694/803 (86%), Positives = 747/803 (92%)

Query: 16  QNSQKHFGYLTERMYSYRDKVLDKKPFIDAERAILVTEAYQKHQEKPNVLKRAYMLQNIL 75
            +    +FG+LT+RM  YR+ VLDKKP+IDAERAIL TEAYQKHQ KP  LKRAYMLQ IL
Sbjct: 3   ETKSPYFGHLTDRMTHYREAVLDKKPYIDAERAILATEAYQKHQNKPANLKRAYMLQTIL 62

Query: 76  EKMTIYIDDETMIVGNQASSDKDAPIFPEYTLEFVVNELDLFEKRDGDVFYITEETKEQI 135
            E MTIYI+DE++I GNQASS+KDAPIFPEYTLEFV+NELDLFEKRDGDVFYITEETK+Q+
Sbjct: 63  ENMTIYIEDESLIAGNQASSNKDAPIFPEYTLEFVLNELDLFEKRDGDVFYITEETKQQL 122

Query: 136 RNIAPFWENNNLRARAGVMLPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLLEEGLIG 195
            R+IAPFWENNNLRAR GV+LPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLLE GL G
Sbjct: 123 RDIAPFWENNNLRARCGVLLPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLLEHGLKG 182

Query: 196 FEKKARKAKADLDLTKPESIDKYHFYDSILITIEAVKTYAERFAILAKKQAKTANAKRRQ 255
            FE++AR AKA LDLT PE+IDKYHFYDS+ I I+AVKTYA+R+A LA++ AKTA  +R+
Sbjct: 183 FEERARAAKAALDLTIPENIDKYHFYDSVFIVIDAVKTYAKRYAKLARELAKTAKPERQA 242

Query: 256 ELLDIASICERVPYYPAETFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMYPYVKSDL 315
            ELLDIA IC++VPY PA+TFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMYPYVK+DL
Sbjct: 243 ELLDIARICDKVPYEPAKTFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMYPYVKADL 302

Query: 316 EAGRETEDSIVERLTNLWIKTITINKVRSQAHTFSSAGSPLYQNVTIGGQTRHKEDAVNP 375
            EAGRETED+IVERLTNLWIKT+TINKVRSQAHTFSSAGSPLYQNVTIGGQTR K+DAVNP
Sbjct: 303 EAGRETEDTIVERLTNLWIKTLTINKVRSQAHTFSSAGSPLYQNVTIGGQTRDKKDAVNP 362

Query: 376 LSFLVLKSVAQTHLPQPNLTVRYHANLDKSFMNEAIEVMKLGFGMPAFNNDEIIIPSFIK 435
            LS+LVL+SVAQT LPQPNLTVRYH  LD +FMNE IEVMKLGFGMPA NNDEIIIPSFIK
Sbjct: 363 LSYLVLRSVAQTKLPQPNLTVRYHKGLDNTFMNECIEVMKLGFGMPAMNNDEIIIPSFIK 422

Query: 436 KGVSEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPKVLLITMNDGIDPASGKRFAPS 495
            KGVSEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPK+LLITMNDGIDPASGKRFA
Sbjct: 423 KGVSEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPKILLITMNDGIDPASGKRFAKG 482

Query: 496 YGHFTQMTSYKELKEAWDKTLRYLTRMSVIVENAIDISLEREVPDILCSALTDDCIGRGK 555
            +GHF  MTSY+ELK AWD TLR +TRMSVIVENAID+ LEREVPDILCSALTDDCIGRGK
Sbjct: 483 HGHFKDMTSYEELKAAWDATLREITRMSVIVENAIDLGLEREVPDILCSALTDDCIGRGK 542

Query: 556 HLKEGGAVYDYISGLQVGIANLSDSLAALKKLVFEEKRLTTLEVWQALQSDYAGPRGEEI 615
             LKEGGAVYDYISGLQVGIANLSDSLAALKKLVFEE RLT  E+W+AL+SD+AG RGE+I
Sbjct: 543 TLKEGGAVYDYISGLQVGIANLSDSLAALKKLVFEEGRLTPEELWKALESDFAGERGEDI 602

Query: 616 RQMLINEAPKYGNDDDYADSLVRECYDVYVEEIAKYPNTRYGRGPIGGIRYSGTSSISAN 675
            RQMLIN+APKYGNDDDYADSLV E YD Y++EIAKYPNTRYGRGPIGGIRYSGTSSISAN
Sbjct: 603 RQMLINDAPKYGNDDDYADSLVVEAYDTYIDEIAKYPNTRYGRGPIGGIRYSGTSSISAN 662

Query: 676 VGQGRGTLATPDGRHAGTPLAEGCSPSHNMDKKGPTSVLKSVSKLPTDEIVGGVLLNQKV 735
            VGQG+GTLATPDGRHAGTPLAEGCSP H+MDKKGPTSVLKSV+KLPTDEIVGGVLLNQKV
Sbjct: 663 VGQGKGTLATPDGRHAGTPLAEGCSPEHSMDKKGPTSVLKSVAKLPTDEIVGGVLLNQKV 722

Query: 736 NPQTLAKEEDKQKLIALLRTFFNRLHGYHIQYNVVSRETLIDAQKHPEKHRDLIVRVAGY 795
            NPQTLAKEEDK KL+ALLRTFFNRLHGYHIQYNVVSRETLIDAQKHPEKHRDLIVRVAGY
Sbjct: 723 NPQTLAKEEDKLKLMALLRTFFNRLHGYHIQYNVVSRETLIDAQKHPEKHRDLIVRVAGY 782

Query: 796 SAFFNVLSKATQDDIIARTEHAL 818
            SAFFNVLSKATQDDII RTEH L
Sbjct: 783 SAFFNVLSKATQDDIIERTEHTL 805
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1001

A DNA sequence (GBSx1061) was identified in *S.agalactiae* <SEQ ID 3071> which encodes the amino acid sequence <SEQ ID 3072>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1024(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA05516 GB:AJ002527 OrfX [Clostridium beijerinckii]
 Identities = 90/214 (42%), Positives = 131/214 (61%), Gaps = 1/214 (0%)

Query: 1   MEFLLDTLNLEAIKKWHHILPLAGVTSNPTIAKKEGDIHFFQRIRDVREIIGREASLHVQ 60
           M+ ++D +N+E IK    I  + GVTSNP+I  K G   +   I+ +RE IG  + LHVQ
Sbjct: 1   MKLIIDDVNIEKIKDVFSIFQIDGVTSNPSILHKYGKQPYEILIK-IREFIGENSELHVQ 59

Query: 61  VVAKDYQGILDDAAKIRQETDDDIYIKVPVTPDGLAAIKTLKAEGYNITATAIYTSMQGL 120
           V+++  +G+L +A KI +E   + Y+K+PVT DGL AIK L+ E  N+TATAIYT MQ
Sbjct: 60  VISESSEGMLKEAHKIIKELGKNTYVKIPVTRDGLKAIKILRKEEINVTATAIYTQMQAY 119

Query: 121 LAISAGADYLAPYFNRMENLDIDATQVIKELAQAIERTGSSSKILAASFKNASQVTKALS 180
           LA  AGA Y APY NR++NL  +  QV K++    E+    +++LAASFKN+ QV +
Sbjct: 120 LAGKAGAQYAAPYVNRIDNLGANGVQVAKDIHDIFEKNNFKTEVLAASFKNSQQVLELCK 179

Query: 181 QGAQSITAGPDIFESVFAMPSIAKAVNDFADDWK 214
            G  + T  PD+ E +     +  AV +F  D++
Sbjct: 180 YGIGAATISPDVIEGLIKNDCVDVAVENFKKDFE 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3073> which encodes the amino acid sequence <SEQ ID 3074>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1090(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 158/222 (71%), Positives = 194/222 (87%)

Query: 1   MEFLLDTLNLEAIKKWHHILPLAGVTSNPTIAKKEGDIHFFQRIRDVREIIGREASLHVQ 60
           ME++LDTL+LEAIKKWHHILPLAGVTSNP+IAKKEG+I FF+RIR+VR IIG +AS+HVQ
Sbjct: 1   MEYMLDTLDLEAIKKWHHILPLAGVTSNPSIAKKEGEIDFFERIREVRAIIGDKASIHVQ 60

Query: 61  VVAKDYQGILDDAAKIRQETDDDIYIKVPVTPDGLAAIKTLKAEGYNITATAIYTSMQGL 120
           V+A+DY+GIL DAA+IR++  D +Y+KVPVT +GLAAIKTLKAEGY+ITATAIYT+ QGL
Sbjct: 61  VIAQDYEGILKDAAEIRRQCGDSVYVKVPVTTEGLAAIKTLKAEGYHITATAIYTTFQGL 120

Query: 121 LAISAGADYLAPYFNRMENLDIDATQVIKELAQAIERTGSSSKILAASFKNASQVTKALS 180
           LAI AGADYLAPY+NRMENL+ID   VI++LA+AI R  ++SKILAASFKN +QV K+ +
Sbjct: 121 LAIEAGADYLAPYYNRMENLNIDPEAVIEQLAEAINRENANSKILAASFKNVAQVNKSFA 180

Query: 181 QGAQSITAGPDIFESVFAMPSIAKAVNDFADDWKASQHSEHI 222
            GAQ+ITAGPD+FE+ FAMPSI KAV+DF  DW+A  H + I
Sbjct: 181 LGAQAITAGPDVFEAGFAMPSIQKAVDDFGKDWEAIHHRKSI 222
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1002

A DNA sequence (GBSx1062) was identified in *S.agalactiae* <SEQ ID 3075> which encodes the amino acid sequence <SEQ ID 3076>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3086(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9545> which encodes amino acid sequence <SEQ ID 9546> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA22477 GB:M65289 glycerol dehydrogenase [Bacillus
           stearothermophilus]
 Identities = 199/362 (54%), Positives = 271/362 (73%), Gaps = 2/362 (0%)

Query: 4   KVFASPSRYIQGKDALFQSIEHIKSLGQTPLILCDDVVYNIVGERFLSYLQD-DLLPHRV 62
           +VF SP++Y+QGK+ + +   +++ +G   +++ D++V+ I G   ++ L+  ++    V
Sbjct: 5   RVFISPAKYVQGKNVITKIANYLEGIGNKTVVIADEIVWKIAGHTIVNELKKGNIAAEEV 64

Query: 63  SFNGEASDNEINRVVAVAKEKNSDLIIGLGGGKTIDSAKAIADKVNLPVVIAPTVASTDA 122
            F+GEAS NE+ R+  +A++  + ++IG+GGGKT+D+AKA+AD+++  +VI PT ASTDA
Sbjct: 65  VFSGEASRNEVERIANIARKAEAAIVIGVGGGKTLDTAKAVADELDAYIVIVPTAASTDA 124

Query: 123 PTSALSVIYTDEGAFEKYIFYSKNPDLVLVDTQVIAQAPKRLLASGIADGLATWVEARAV 182
           PTSALSVIY+D+G FE Y FY KNPDLVLVDT++IA AP RLLASGIAD LATWVEAR+V
Sbjct: 125 PTSALSVIYSDDGVFESYRFYKKNPDLVLVDTKIIANAPPRLLASGIADALATWVEARSV 184

Query: 183 LQKNGIAMAGGRQTLAGVAIAQACERTLFNDSLQALAACDAKVVTKALENVIEANTLLSG 242
           ++  G  MAGG  T+A  AIA+ CE+TLF     A  +  AKVVT ALE V+EANTLLSG
Sbjct: 185 IKSGGKTMAGGIPTIAAEAIAEKCEQTLFKYGKLAYESVKAKVVTPALEAVVEANTLLSG 244

Query: 243 LGFESAGLAAAHAIHNGFTALSGDIHHLTHGEKVAYGTLTQLFLENRPKEEIDRYINLYQ 302
           LGFES GLAAAHAIHNGFTAL G+IHHLTHGEKVA+GTL QL LE   ++EI+RYI LY
Sbjct: 245 LGFESGGLAAAHAIHNGFTALEGEIHHLTHGEKVAFGTLVQLALEEHSQQEIERYIELYL 304

Query: 303 AIGMPTTLAELHLGDATYEELLKVGQQATIEGETIHEMPFKISAEDVAAALLTVDRYVSN 362
           ++ +P TL ++ L DA+ E++LKV + AT EGETIH   F ++A+DVA A+   D+Y
Sbjct: 305 SLDLPVTLEDIKLKDASREDILKVAKAATAEGETIHN-AFNVTADDVADAIFAADQYAKA 363

Query: 363 HQ 364
           ++
Sbjct: 364 YK 365
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3077> which encodes the amino acid sequence <SEQ ID 3078>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.62   Transmembrane  101 - 117 (  98 - 119)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA22477 GB:M65289 glycerol dehydrogenase [Bacillus
           stearothermophilus]
 Identities = 202/357 (56%), Positives = 261/357 (72%), Gaps = 1/357 .(0%)

Query: 2   KVFASPSRYIQGKNALFTNVKTLKQLGDSPILLCDDVVYGIVGERFESYLIDNGMTPVHV 61
           +VF SP++Y+QGKN +      L+ +G+  +++ D++V+ I G    + L    +    V
Sbjct: 5   RVFISPAKYVQGKNVITKIANYLEGIGNKTVVIADEIVWKIAGHTIVNELKKGNIAAEEV 64

Query: 62  AFNGEASDNEISRVVAIAKENGNDVIIGLGGGKTIDSAKAIADLLAVPVIIAPTIASTDA 121
            F+GEAS NE+ R+  IA++    ++IG+GGGKT+D+AKA+AD L   ++I PT ASTDA
Sbjct: 65  VFSGEASRNEVERIANIARKAEAAIVIGVGGGKTLDTAKAVADELDAYIVIVPTAASTDA 124

Query: 122 PTSALSVIYTDEGAFEKYIFYSKNPDLVLVDTQVICQAPKRLLASGIADGLATWVEARAV 181
           PTSALSVIY+D+G FE Y FY KNPDLVLVDT++I  AP RLLASGIAD LATWVEAR+V
Sbjct: 125 PTSALSVIYSDDGVFESYRFYKKNPDLVLVDTKIIANAPPRLLASGIADALATWVEARSV 184

Query: 182 MQKNGDTMAGGNQTLAGVAIAKACEQTLFADGLKAMASCDRQVVTPALENVIEANTLLSG 241
           ++  G TMAGG  T+A  AIA+ CEQTLF  G  A  S   +VVTPALE V+EANTLLSG
Sbjct: 185 IKSGGKTMAGGIPTIAAEAIAEKCEQTLFKYGKLAYESVKAKVVTPALEAVVEANTLLSG 244

Query: 242 LGFESAGLAAAHAIHNGFTALTGAIHHLTHGEKVAYGTLTQLFLENRSREEIDRYIDFYQ 301
           LGFES GLAAAHAIHNGFTAL G IHHLTHGEKVA+GTL QL LE  S++EI+RYI+ Y
Sbjct: 245 LGFESGGLAAAHAIHNGFTALEGEIHHLTHGEKVAFGTLVQLALEEHSQQEIERYIELYL 304

Query: 302 AIGMPTTLKEMHLDTATQEDFLKIGRQATMAGETIHQMPFVISPEDVAAALVAVDAY 358
           ++ +P TL+++ L  A++ED LK+ + AT  GETIH   F ++ +DVA A+ A D Y
Sbjct: 305 SLDLPVTLEDIKLKDASREDILKVAKAATAEGETIHN-AFNVTADDVADAIFAADQY 360
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 287/361 (79%), Positives = 325/361 (89%), Gaps = 1/361 (0%)

Query: 3   MKVFASPSRYIQGKDALFQSIEHIKSLGQTPLILCDDVVYNIVGERFLSYLQDD-LLPHR 61
           MKVFASPSRYIQGK+ALF +++ +K LG +P++LCDDVVY IVGERF SYL D+ + P
Sbjct: 1   MKVFASPSRYIQGKNALFTNVKTLKQLGDSPILLCDDVVYGIVGERFESYLIDNGMTPVH 60

Query: 62  VSFNGEASDNEINRVVAVAKEKNSDLIIGLGGGKTIDSAKAIADKVNLPVVIAPTVASTD 121
           V+FNGEASDNEI+RVVA+AKE  +D+IIGLGGGKTIDSAKAIAD + +PV+IAPT+ASTD
Sbjct: 61  VAFNGEASDNEISRVVAIAKENGNDVIIGLGGGKTIDSAKAIADLLAVPVIIAPTIASTD 120

Query: 122 APTSALSVIYTDEGAFEKYIFYSKNPDLVLVDTQVIAQAPKRLLASGIADGLATWVEARA 181
           APTSALSVIYTDEGAFEKYIFYSKNPDLVLVDTQVI QAPKRLLASGIADGLATWVEARA
Sbjct: 121 APTSALSVIYTDEGAFEKYIFYSKNPDLVLVDTQVICQAPKRLLASGIADGLATWVEARA 180

Query: 182 VLQKNGIAMAGGRQTLAGVAIAQACERTLFNDSLQALAACDAKVVTKALENVIEANTLLS 241
           V+QKNG  MAGG QTLAGVAIA+ACE+TLF D L+A+A+CD +VVT ALENVIEANTLLS
Sbjct: 181 VMQKNGDTMAGGNQTLAGVAIAKACEQTLFADGLKAMASCDRQVVTPALENVIEANTLLS 240

Query: 242 GLGFESAGLAAAHAIHNGFTALSGDIHHLTHGEKVAYGTLTQLFLENRPKEEIDRYINLY 301
           GLGFESAGLAAAHAIHNGFTAL+G IHHLTHGEKVAYGTLTQLFLENR +EEIDRYI+ Y
Sbjct: 241 GLGFESAGLAAAHAIHNGFTALTGAIHHLTHGEKVAYGTLTQLFLENRSREEIDRYIDFY 300

Query: 302 QAIGMPTTLAELHLGDATYEELLKVGQQATIEGETIHEMPFKISAEDVAAALLTVDRYVSN 362
           QAIGMPTTL E+HL  AT E+ LK+G+QAT+ GETIH+MPF IS EDVAAAL+ VD YV++
Sbjct: 301 QAIGMPTTLKEMHLDTATQEDFLKIGRQATMAGETIHQMPFVISPEDVAAALVAVDAYVTS 361
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1003

A DNA sequence (GBSx1063) was identified in *S.agalactiae* <SEQ ID 3079> which encodes the amino acid sequence <SEQ ID 3080>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  262 - 278 ( 262 - 279)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA88310 GB:AB028865 O-acetylserine lyase [Streptococcus suis]
 Identities = 239/304 (78%), Positives = 273/304 (89%)

Query: 4   IYNSITDLIGNTPIIQLHHIVPEGAAEVYVKLESFNPGSSVKDRIALAMIEDAEQKGILK 63
           IY +IT L+G TP+I+L++IVPEGAAEVYVKLE+FNPGSSVKDRIALAMIEDAE+ G +K
Sbjct: 3   IYQNITQLVGKTPVIKLNNIVPEGAAEVYVKLEAFNPGSSVKDRIALAMIEDAEKAGTIK 62

Query: 64  AGDTIVEPTSGNTGIGLAWVGKAKGYNVIIVMPETMSIERRKIIQAYGAQLVLTPGSEGM 123
            GDTIVEPTSGNTGIGLAWVG AKGYNVIIVMPETMS+ERRKIIQAYGA+LVLTPGSEGM
Sbjct: 63  PGDTIVEPTSGNTGIGLAWVGAAKGYNVIIVMPETMSVERRKIIQAYGAELVLTPGSEGM 122

Query: 124 KGAIAKAKEISAEQNAWLPLQFNNQANPEIHEKTTGREIIETFGEKGLDAFIAGVGTGGT 183
           KGAIAKAKEI+ E+N W+P QF N +NP++HE TTG+EI+E FG  GLDAF++GVGTGGT
Sbjct: 123 KGAIAKAKEIAEEKNGWVPFQFANPSNPKVHEDTTGQEILEDFGTTGLDAFVSGVGTGGT 182

Query: 184 ITGVSRALKKVNPDVAIYAVEADESAILSGEQPGPHKIQGISAGFIPETLATDSYDHIIR 243
           ++GVS  LK  NPD+AIYAVEADESA+LSGE PGPHKIQGISAGFIP+TL T +YD IIR
Sbjct: 183 VSGVSHVLKTANPDIAIYAVEADESAVLSGEAPGPHKIQGISAGFIPDTLDTSAYDGIIR 242

Query: 244 VTSDDAIETGRIIGGLEGFLAGISASAAIYAAIEVAKQLGKGKKVLALLPDNGERYLSTS 303
           V SDDA+ TGR IGG EGFL GIS+ AAI+AAIEVAK+LG GKKVLA+LPDNGERYLST+
Sbjct: 243 VKSDDALATGRAIGGKEGFLVGISSGAAIHAAIEVAKELGTGKKVLAILPDNGERYLSTA 302

Query: 304 LYDF 307
           LY+F
Sbjct: 303 LYEF 306
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3081> which encodes the amino acid sequence <SEQ ID 3082>. Analysis of this protein sequence reveals the following:

```
     Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.48   Transmembrane  262 - 278 ( 262 - 278)

----- Final Results -----
               bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA88310 GB:AB028865 O-acetylserine lyase [Streptococcus suis]
 Identities = 235/303 (77%), Positives = 261/303 (85%)

Query: 4   IYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLIS 63
           IY+ IT+LVG+TP+IKLN ++P  AA+VYVKLEAFNPGSSVKDRIAL+MIE AE  G I
Sbjct: 3   IYQNITQLVGKTPVIKLNNIVPEGAAEVYVKLEAFNPGSSVKDRIALAMIEDAEKAGTIK 62

Query: 64  PGDVIIEPTSGNTGIGLAWVGAAKGYRVIIVMPETMSLERRQIIQAYGAELVLTPGAEGM 123
           PGD I+EPTSGNTGIGLAWVGAAKGY VIIVMPETMS+ERR+IIQAYGAELVLTPG+EGM
```

```
Sbjct: 63  PGDTIVEPTSGNTGIGLAWVGAAKGYNVIIVMPETMSVERRKIIQAYGAELVLTPGSEGM 122

Query: 124 KGAIAKAETLAIELGAWMPMQFNNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGT 183
           KGAIAKA+ +A E   W+P QF NP+NP +HE TT QEILE F    LDAFVSGVGTGGT
Sbjct: 123 KGAIAKAKEIAEEKNGWVPFQFANPSNPKVHEDTTGQEILEDFGTTGLDAFVSGVGTGGT 182

Query: 184 LSGVSHVLKKANPETVIYAVEAEESAVLSGQEPGPHKIQGISAGFIPNTLDTKAYDQIIR 243
           +SGVSHVLK ANP+  IYAVEA+ESAVLSG+ PGPHKIQGISAGFIP+TLDT AYD IIR
Sbjct: 183 VSGVSHVLKTANPDIAIYAVEADESAVLSGEAPGPHKIQGISAGFIPDTLDTSAYDGIIR 242

Query: 244 VKSKDALETARLTGAKEGFLVGISSGAALYAAIEVAKQLGKGKHVLTILPDNGERYLSTE 303
           VKS DAL T R  G KEGFLVGISSGAA++AAIEVAK+LG GK VL ILPDNGERYLST
Sbjct: 243 VKSDDALATGRAIGGKEGFLVGISSGAAIHAAIEVAKELGTGKKVLAILPDNGERYLSTA 302

Query: 304 LYD 306
           LY+
Sbjct: 303 LYE 305
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 222/306 (72%), Positives = 263/306 (85%)

Query: 1   MSKIYNSITDLIGNTPIIQLHHIVPEGAAEVYVKLESFNPGSSVKDRIALAMIEDAEQKG 60
           M+KIY +IT+L+G TPII+L+ ++P  AA+VYVKLE+FNPGSSVKDRIAL+MIE AE +G
Sbjct: 1   MTKIYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEG 60

Query: 61  ILKAGDTIVEPTSGNTGIGLAWVGKAKGYNVIIVMPETMSIERRKIIQAYGAQLVLTPGS 120
           ++  GD I+EPTSGNTGIGLAWVG AKGY VIIVMPETMS+ERR+IIQAYGA+LVLTPG+
Sbjct: 61  LISPGDVIIEPTSGNTGIGLAWVGAAKGYRVIIVMPETMSLERRQIIQAYGAELVLTPGA 120

Query: 121 EGMKGAIAKAKEISAEQNAWLPLQFNNQANPEIHEKTTGREIIETFGEKGLDAFIAGVGT 180
           EGMKGAIAKA+ ++ E  AW+P+QFNN ANP IHEKTT +EI+E F E  LDAF++GVGT
Sbjct: 121 EGMKGAIAKAETLAIELGAWMPMQFNNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGT 180

Query: 181 GGTITGVSRALKKVNPDVAIYAVEADESAILSGEQPGPHKIQGISAGFIPETLATDSYDH 240
           GGT++GVS  LKK NP+  IYAVEA+ESA+LSG++PGPHKIQGISAGFIP TL T +YD
Sbjct: 181 GGTLSGVSHVLKKANPETVIYAVEAEESAVLSGQEPGPHKIQGISAGFIPNTLDTKAYDQ 240

Query: 241 IIRVTSDDAIETGRIIGGLEGFLAGISASAAIYAAIEVAKQLGKGKKVLALLPDNGERYL 300
           IIRV S DA+ET R+ G  EGFL GIS+ AA+YAAIEVAKQLGKGK VL +LPDNGERYL
Sbjct: 241 IIRVKSKDALETARLTGAKEGFLVGISSGAALYAAIEVAKQLGKGKHVLTILPDNGERYL 300

Query: 301 STSLYD 306
           ST LYD
Sbjct: 301 STELYD 306
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1004

A DNA sequence (GBSx1064) was identified in *S.agalactiae* <SEQ ID 3083> which encodes the amino acid sequence <SEQ ID 3084>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3666(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07349 GB:AP001519 unknown conserved protein [Bacillus halodurans]
 Identities = 96/204 (47%), Positives = 127/204 (62%)
```

```
Query: 2   NYKTIKSDGIVEEEIKKSRFICHLKRVESEEEGRNYITQIKKAHYKANHSCSAMVIGEKG 61
           +Y T+K  GI E  I+KSRFI HL R  SEEE   +I QIKK H+ A H+CSA +IGE
Sbjct: 4   SYYTVKESGIHEISIQKSRFIAHLSRATSEEEAIQFIEQIKKEHWNATHNCSAYLIGEND 63

Query: 62  DIKRSSDDGEPSGTAGIPMLTVLEKQGLTNVVAVVTRYFGGIKLGAGGLIRAYSGSVANT 121
            +++++DDGEPSGTAG+PML VL+K+ L + VAVVTRYFGG+KLGAGGLIRAY  +V++
Sbjct: 64  QVQKANDDGEPSGTAGVPMLEVLKKRRLKDTVAVVTRYFGGVKLGAGGLIRAYGSAVSDG 123

Query: 122 IKEIGVVEVKEQIGIRIQLTYPQYQTFDNFLKEHHLQEFETEFLEAVTCKIYVDPKEFEH 181
           +  IGVVE K    I   + Y      +N L++ H    E  +LE V  + YV   E E
Sbjct: 124 LNAIGVVERKRMQVIHTSIDYHWLGKVENELRQSHYLLKEISYLENVDVQTYVLEAEVES 183

Query: 182 TITNLTEFYQGKALLTEEGSQIVE 205
               +T    G+A  T    + +E
Sbjct: 184 YCEWMTNLTNGQAAFTHGAIEYLE 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3085> which encodes the amino acid sequence <SEQ ID 3086>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43    Transmembrane   86 - 102 (  86 - 102)

----- Final Results -----
              bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9153> which encodes the amino acid sequence <SEQ ID 9154>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43    Transmembrane   81 -  97 (  81 -  97)

----- Final Results -----
              bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 122/206 (59%), Positives = 153/206 (74%)

Query: 2   NYKTIKSDGIVEEEIKKSRFICHLKRVESEEEGRNYITQIKKAHYKANHSCSAMVIGEKG 61
           ++KTIK+ G  EE IKKSRFICH+KRV +EE+G+N++  IKK HYKANHSC AM+IG
Sbjct: 8   HFKTIKASGFFEESIKKSRFICHIKRVSTEEDGKNFVNAIKKEHYKANHSCFAMIIGNNR 67

Query: 62  DIKRSSDDGEPSGTAGIPMLTVLEKQGLTNVVAVVTRYFGGIKLGAGGLIRAYSGSVANT 121
            IKRSSDDGEPSGTAGIP+L+VLEKQ LTNVV VVTRYFGGIKLG GGLIRAYS   A
Sbjct: 68  QIKRSSDDGEPSGTAGIPILSVLEKQCLTNVVVVVTRYFGGIKLGTGGLIRAYSNMTATA 127

Query: 122 IKEIGVVEVKEQIGIRIQLTYPQYQTFDNFLKEHHLQEFETEFLEAVTCKIYVDPKEFEH 181
           IK  G++EVK+QIG+ I L+YPQYQ + N L +  L E ET+F + +   +Y D +  E+
Sbjct: 128 IKRFGIIEVKQQIGLEITLSYPQYQLYSNLLDQLALTETETKFSDTIKTTLYCDTERVEN 187

Query: 182 TITNLTEFYQGKALLTEEGSQIVEIP 207
            I  LT +Y G+    + GS+++E P
Sbjct: 188 LIDTLTNYYHGQISCEKIGSKVIEFP 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1005

A DNA sequence (GBSx1065) was identified in *S.agalactiae* <SEQ ID 3087> which encodes the amino acid sequence <SEQ ID 3088>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44940 GB:U56901 involved in transformation [Bacillus subtilis]
 Identities = 160/405 (39%), Positives = 228/405 (55%), Gaps = 20/405 (4%)

Query: 35  YICTRCSSSVAKNCQL----PTGNYYCRECIVFGRVTSNENLYYFPQKTFSKTNSLK--W 88
            Y C RC     +  +          YCR C++ GRV+    LY + ++  S   S+K  W
Sbjct: 58  YRCNRCGQTDQRYFSFYHSSGKNKLYCRSCVMMGRVSEEVPLYSWKEENESNWKSIKLTW 117

Query: 89  KGELTPYQNEVSEELLKGISSKENLLVHAVTGAGKTEMIYHSVAKVIDTGGSVCIASPRI 148
             G+L+  Q + +  L++ IS KE LL+ AV GAGKTEM++  +   ++ G  VCIA+PR
Sbjct: 118 DGKLSSGQQKAANVLIEAISKKEELLIWAVCGAGKTEMLFPGIESALNQGLRVCIATPRT 177

Query: 149 DVCLELYKRLSNDFRCA-ITLMHGESPSYQR-SPLTIATTHQLLKFYHAFDLLIVDEVDA 206
            DV LEL  RL   F+ A I+ ++G S    R SPL I+TTHQLL++  A D++I+DEVDA
Sbjct: 178 DVVLELAPRLKAAFQGADISALYGGSDDKGRLSPLMISTTHQLLRYKDAIDVMIIDEVDA 237

Query: 207 FPYVDNPILYQGVKQALKENGTSIFLTATSTTELERKVARKELKKLHLARRFHANPLVIP 266
            FPY  +  L   V++A K+N T ++L+AT   EL+RK    +L  + +  R H  PL  P
Sbjct: 238 FPYSADQTLQFAVQKARKKNSTLVYLSATPPKELKRKALNGQLHSVRIPARHHRKPLPEP 297

Query: 267 EMVWVSGIQKSLQTQKLPPKLYQLINKQRQTRYPLLLFFPHISEGQVFTEILRQAFPMEK 326
              VW    +K L   K+PP + + I    +   P+ LF P +S       IL +A    K
Sbjct: 298 RFVWCGNWKKKLNRNKIPPAVKRWIEFHVKEGRPVFLFVPSVS-------ILEKAAACFK 350

Query: 327 -----IGFVSSKSTSRLKLVQDFRDNKLSILVSTTILERGVTFPSVDVFVIQANHHLFTK 381
                    V ++   R + VQ FRD +L +L++TTILERGVT P V   V+ A   +FT+
Sbjct: 351 GVHCRTASVHAEDKHRKEKVQQFRDGQLDLLITTTILERGVTVPKVQTGVLGAESSIFTE 410

Query: 382 SSLVQISGRVGRALERPEGLLYFLHDGKSKSMHQAIKEIKNMNHI 426
            S+LVQI+GR GR  E  +G + + H GK+KSM  A K IK MN +
Sbjct: 411 SALVQIAGRTGRHKEYADGDVIYFHFGKTKSMLDARKHIKEMNEL 455
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3089> which encodes the amino acid sequence <SEQ ID 3090>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.09   Transmembrane  304 - 320 ( 303 - 322)

----- Final Results -----
               bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:U56901 involved in transformation [Bacillus subt...    258  1e-67

>GP:AAC44940 GB:U56901 involved in transformation [Bacillus subtilis]
 Identities = 155/435 (35%), Positives = 249/435 (56%), Gaps = 20/435 (4%)

Query: 10  RLLLESQLPDSAKQLAQPLK--------SVVILRGKMICQRCHYQLDEEA-----RLPSG 56
```

```
           R LL ++L  S + +   +K        S+ I + +  C RC  Q D+
Sbjct: 22  RHLLRTELSFSDEMIEWHIKNGYITAENSISINKRRYRCNRCG-QTDQRYFSFYHSSGKN 80

Query: 57  AYYCRFCLVFGRNQSDKLLYAIPPMHFP--KGNYLVWGGQLTAYQEMISQQLLINMQNQK 114
             YCR C++ GR   +  LY+    +    K   L W G+L++ Q+  +  L+  +  ++
Sbjct: 81  KLYCRSCVMMGRVSEEVPLYSWKEENESNWKSIKLTWDGKLSSGQQKAANVLIEAISKKE 140

Query: 115 TTLVHAVTGAGKTEMIYAAIEAVINTGGWVCIASPRVDVCVEVATRLSQAFS-CSICLMH 173
             L+ AV GAGKTEM++  IE+ +N G  VCIA+PR DV +E+A RL  AF    I  ++
Sbjct: 141 ELLIWAVCGAGKTEMLFPGIESALNQGLRVCIATPRTDVVLELAPRLKAAFQGADISALY 200

Query: 174 AESLPYQR-APIIVATTHQLLKFHKAFDLLIIDEVDAFPFVNNIQLHYAASQALKEGGAK 232
             S    R +P++++TTHQLL++  A D++IIDEVDAFP+  +  L +A  +A K+
Sbjct: 201 GGSDDKGRLSPLMISTTHQLLRYKDAIDVMIIDEVDAFPYSADQTLQFAVQKARKKNSTL 260

Query: 233 ILLTATSTRTLERKVNKGEVVKLTLARRFHNRPLVIPKFIRSFNLFKMIHRQKLPLKILK 292
           + L+AT  + L+RK   G++  + +  R H +PL  P+F+   N  K ++R K+P  + +
Sbjct: 261 VYLSATPPKELKRKALNGQLHSVRIPARHHRKPLPEPRFVWCGNWKKKLNRNKIPPAVKR 320

Query: 293 YLKKQRKTGYPLLIFLPTIIMAESVTAILKELLPAEQIACVSSQSQNRKEDITAFRQGKK 352
           +++   K G P+ +F+P++ + E   A  K +    + A V ++ ++RKE +  FR G+
Sbjct: 321 WIEFHVKEGRPVFLFVPSVSILEKAAACFKGV--HCRTASVHAEDKHRKEKVQQFRDGQL 378

Query: 353 TILITTSILERGVTFPQIDVFVLGSHHRVYSSQSLVQIAGRVGRSIDRPDGTLYFFHEGI 412
            +LITT+ILERGVT P++   VLG+   +++  +LVQIAGR GR  +  DG + +FH G
Sbjct: 379 DLLITTTILERGVTVPKVQTGVLGAESSIFTESALVQIAGRTGRHKEYADGDVIYFHFGK 438

Query: 413 SKAMLLARKEIKEMN 427
           +K+ML ARK IKEMN
Sbjct: 439 TKSMLDARKHIKEMN 453
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 223/427 (52%), Positives = 299/427 (69%)

Query: 1   MENYLGRLWTKAQLSEQLRKIAISLPSFIKKGSDYICTRCSSSVAKNCQLPTGNYYCREC 60
           +EN  GRL  ++QL +  +++A  L S +      IC RC   + +  +LP+G YYCR C
Sbjct: 4   IENSYGRLLLESQLPDSAKQLAQPLKSVVILRGKMICQRCHYQLDEEARLPSGAYYCRFC 63

Query: 61  IVFGRVTSNENLYYFPQKTFSKTNSLKWKGELTPYQNEVSEELLKGISSKENLLVHAVTG 120
           +VFGR  S++ LY  P   F K N L W G+LT YQ  +S++LL  + +++  LVHAVTG
Sbjct: 64  LVFGRNQSDKLLYAIPPMHFPKGNYLVWGGQLTAYQEMISQQLLINMQNQKTTLVHAVTG 123

Query: 121 AGKTEMIYHSVAKVIDTGGSVCIASPRIDVCLELYKRLSNDFRCAITLMHGESPSYQRSP 180
           AGKTEMIY ++  VI+TGG VCIASPR+DVC+E+  RLS  F C+I LMH ES  YQR+P
Sbjct: 124 AGKTEMIYAAIEAVINTGGWVCIASPRVDVCVEVATRLSQAFSCSICLMHAESLPYQRAP 183

Query: 181 LTIATTHQLLKFYHAFDLLIVDEVDAFPYVDNPILYQGVKQALKENGTSIFLTATSTTEL 240
           + +ATTHQLLKF+ AFDLLI+DEVDAFP+V+N  L+    QALKE G  I LTATST  L
Sbjct: 184 IIVATTHQLLKFHKAFDLLIIDEVDAFPFVNNIQLHYAASQALKEGGAKILLTATSTRTL 243

Query: 241 ERKVARKELKKLHLARRFHANPLVIPEMVWVSGIQKSLQTQKLPPKLYQLINKQRQTRYP 300
           ERKV + E+ KL LARRFH  PLVIP+ +    + K +  QKLP K+ + + KQR+T YP
Sbjct: 244 ERKVNKGEVVKLTLARRFHNRPLVIPKFIRSFNLFKMIHRQKLPLKILKYLKKQRKTGYP 303

Query: 301 LLLFFPHISEGQVFTEILRQAFPMEKIGFVSSKSTSRLKLVQDFRDNKLSILVSTTILER 360
           LL+F P I   +  T IL++  P E+I  VSS+S +R + +  FR  K +IL++T+ILER
Sbjct: 304 LLIFLPTIIMAESVTAILKELLPAEQIACVSSQSQNRKEDITAFRQGKKTILITTSILER 363

Query: 361 GVTFPSVDVFVIQANHHLFTKSSLVQISGRVGRALERPEGLLYFLHDGKSKSMHQAIKEI 420
           GVTFP +DVFV+ ++H +++  SLVQI+GRVGR+++RP+G LYF H+G SK+M  A KEI
Sbjct: 364 GVTFPQIDVFVLGSHHRVYSSQSLVQIAGRVGRSIDRPDGTLYFFHEGISKAMLLARKEI 423

Query: 421 KNMNHIG 427
           K MN+ G
Sbjct: 424 KEMNYKG 430
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1006

A DNA sequence (GBSx1066) was identified in *S.agalactiae* <SEQ ID 3091> which encodes the amino acid sequence <SEQ ID 3092>. This protein is predicted to be comf operon protein 3 (comFC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0894(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44942 GB:U56901 involved in transformation [Bacillus subtilis]
 Identities = 76/230 (33%), Positives = 118/230 (51%), Gaps = 11/230 (4%)

Query: 1   MTCLLCHEIDLSQLTFVELMLLKPKQNVICQTCKGSFEALSREMGCQTCCK-QIPQKQCQ 59
           M CLLC       +T+  L LLKP + V C +C+   + ++  + C  C + Q     C+
Sbjct: 1   MICLLCDSQFSQDVTWRALFLLKPDEKV-CYSCRSKLKKITGHI-CPLCGRPQSVHAVCR 58

Query: 60  DCIYWGKKGIEV----NHFSLYRYNEAMKKNFSLFKFQGDYLLKDVFTKEIKAALKKY-- 113
           DC  W  +  +      + S+Y YN+ MK+  S FKF+GD  + + F  +  +   K
Sbjct: 59  DCEVWRTRIRDSLLLRQNRSVYTYNDMMKETLSRFKFRGDAEIINAFKSDFSSTFSKVYP 118

Query: 114 -KGYTIVPVPLSHEGYQNRQFNQVIAFLQSANIPYKNILSKKDGGKQSANNKEERLKQVQ 172
            K + +VP+PLS E  + R FNQ     +  + P  + L + +  KQS   K ERL
Sbjct: 119 DKHFVLVPIPLSKEREEERGFNQAHLLAECLDRPSHHPLIRLNNEKQSKKKKTERLLSEC 178

Query: 173 QFTLKNEAELGDNLLIVDDIYTTGATIAQIRKLLEEKG-IKNIKSFSLAR 221
            F  KN +  G N++++DD+YTTGAT+    + L EKG   ++ SF+L R
Sbjct: 179 IFDTKNNSAEGMNIILIDDLYTTGATLHFAARCLLEKGKAASVSSFTLIR 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3093> which encodes the amino acid sequence <SEQ ID 3094>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0763(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 100/222 (45%), Positives = 139/222 (62%), Gaps = 2/222 (0%)

Query: 1   MTCLLCHEIDLSQLTFVELMLLKPKQNVICQTCKGSFEALSREMGCQTCCKQIPQKQCQD 60
           M CLLC +I  + ++  E++ L+   + ICQ C+ SF+ + + + C TCC       C+D
Sbjct: 1   MICLLCQQISQTPISITEIIFLRRISSPICQQCQKSFQKIGKSV-CATCCANSDIIACRD 59

Query: 61  CIYWGKKGIEVNHFSLYRYNEAMKKNFSLFKFQGDYLLKDVFTKEIKAALKKY-KGYTIV 119
           C+ W  KG  VNH SLY YN AMK  FS +KFQGDYLL+ VF  E+   + KY KGY  V
Sbjct: 60  CLKWENKGYNVNHRSLYCYNAAMKAYFSQYKFQGDYLLRKVFAVELADVITKYYKGYIPV 119

Query: 120 PVPLSHEGYQNRQFNQVIAFLQSANIPYKNILSKKDGGKQSANNKEERLKQVQQFTLKNE 179
           PVP+S   ++ RQFNQV A L++AN+ Y ++  K D   QS+  K+ERL   + + L
Sbjct: 120 PVPVSPGCFRERQFNQVSAILEAANVSYLSLFEKLDNTHQSSRTKKERLLVEKSYRLLKV 179
```

```
Query: 180 AELGDNLLIVDDIYTTGATIAQIRKLLEEKGIKNIKSFSLAR 221
           + + D +LIVDDIYTTG+TI  +RK L +    +IKS S+AR
Sbjct: 180 SNIPDKILIVDDIYTTGSTIIALRKQLAKVANSDIKSLSIAR 221
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1007

A DNA sequence (GBSx1067) was identified in *S.agalactiae* <SEQ ID 3095> which encodes the amino acid sequence <SEQ ID 3096>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3889(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB91549 GB:AJ249134 hypothetical protein [Lactococcus lactis]
 Identities = 107/185 (57%), Positives = 140/185 (74%), Gaps = 3/185 (1%)

Query: 1   MIKYSIRGENIEVTEAIREYVETKLSKVEKYFNEAQELDTRVNLKVYREKTAKVEVTILI 60
            MIK++IRGEN+EVT+AIR YVE K+ K++KYFN+  E+   VNLKVY EK AKVEVT+
Sbjct: 1   MIKFNIRGENVEVTDAIRAYVEDKIGKLDKYFNDGHEVTAYVNLKVYTEKRAKVEVTLPA 60

Query: 61  DSITLRAEDVSQDMYGSIDLVVDKIERQIRKNKTKIAKKYREKIPASQVFTTEFEAEPDE 120
             ++TLRAED SQDMY SID V +K+ERQIRK KT++ +K R  +P  QVF  EF
Sbjct: 61  KNVTLRAEDTSQDMYSSIDFVEEKLERQIRKYKTRMNRKPRNAVPTGQVFGDEFAPLDTT 120

Query: 121 EAVSQ---RIVRTKNVNLKPMDVEEALLQMELLGHDFFIYTDAEDNTTNVLYKREDGELG 177
            + V++    IVRTK+V LKPMD EEA+LQM++LGHDF+++TDA+ N T+V+Y+R DG  G
Sbjct: 121 DEVAEDHVDIVRTKHVALKPMDAEEAVLQMDMLGHDFYVFTDADSNGTHVVYRRTDGRYG 180

Query: 178 LIEAK 182
            LIE +
Sbjct: 181 LIETE 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3097> which encodes the amino acid sequence <SEQ ID 3098>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 145/182 (79%), Positives = 165/182 (89%)

Query: 1   MIKYSIRGENIEVTEAIREYVETKLSKVEKYFNEAQELDTRVNLKVYREKTAKVEVTILI 60
            MIK+SIRGENIEVTEAIR+YVE+KL+K+EKYF + QE+D RVNLKVYRE+++KVEVTI +
Sbjct: 1   MIKFSIRGENIEVTEAIRDYVESKLTKIEKYFAKDQEIDARVNLKVYRERSSKVEVTIPL 60

Query: 61  DSITLRAEDVSQDMYGSIDLVVDKIERQIRKNKTKIAKKYREKIPASQVFTTEFEAEPDE 120
            DS+TLRAEDVSQDMYGSIDLVVDKIERQIRKNKTKIAKK+REK+P  QVFTTEFEAE +
Sbjct: 61  DSVTLRAEDVSQDMYGSIDLVVDKIERQIRKNKTKIAKKHREKVPTGQVFTTEFEAEEVD 120
```

```
Query: 121 EAVSQRIVRTKNVNLKPMDVEEALLQMELLGHDFFIYTDAEDNTTNVLYKREDGELGLIE 180
           E    ++VRTKNV LKPMDVEEA LQMELLGHDFFIYTD+ED  TN+LY+REDG LGLIE
Sbjct: 121 EIPEVQVVRTKNVTLKPMDVEEARLQMELLGHDFFIYTDSEDGATNILYRREDGNLGLIE 180

Query: 181 AK 182
           AK
Sbjct: 181 AK 182
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1008

A DNA sequence (GBSx1068) was identified in *S.agalactiae* <SEQ ID 3099> which encodes the amino acid sequence <SEQ ID 3100>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0685(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1009

A DNA sequence (GBSx1077) was identified in *S.agalactiae* <SEQ ID 3101> which encodes the amino acid sequence <SEQ ID 3102> (sgaT). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.95   Transmembrane   99 - 115 (  87 - 115)
   INTEGRAL    Likelihood = -3.50   Transmembrane   43 -  59 (  42 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03942 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 47/111 (42%), Positives = 76/111 (68%), Gaps = 5/111 (4%)

Query: 1   MAIIYLIVAVFAG--EAYIAKEI---SNGVNGLVYALQLAGQFAAGVFVILAGVRLILGE 55
           M I++L+ A+     +   A+E+   S   +  +YA+  +  FA G+ V+L GV++ +GE
Sbjct: 233 MGILFLVGAIILALKDTQGAQELIAQSGEQSFFIYAIIQSFMFAGGIAVVLLGVKMFIGE 292

Query: 56  IVPAFKGISEKLVPNSKPALDCPIVYPYAPNAVLIGFISKFVGGLVSMIVM 106
           +VPAF GI+ KLVP ++PALD P+V+P APNAV++GF+  FVG L+ ++V+
Sbjct: 293 VVPAFNGIATKLVPGARPALDAPVVFPMAPNAVILGFLGAFVGALIWLVVI 343
```

There is also homology to SEQ ID 516.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1010

A DNA sequence (GBSx1078) was identified in *S.agalactiae* <SEQ ID 3103> which encodes the amino acid sequence <SEQ ID 3104>. This protein is predicted to be tryptophanyl-tRNA synthetase (trpS). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2156(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05711 GB:L49336 tryptophanyl-tRNA synthetase [Clostridium
           longisporum]
 Identities = 225/340 (66%), Positives = 271/340 (79%), Gaps = 3/340 (0%)

Query: 1   MTKPIILTGDRPTGKLHIGHYVGSLKNRVLLQNEGSYTLFVFLADQQALTDHAKDPQTIV 60
           M K IILTGDRPTGKLHIGHYVGSLKNRV LQN G Y  F+ +ADQQALTD+A++P+ I
Sbjct: 1   MAKEIILTGDRPTGKLHIGHYVGSLKNRVQLQNSGDYRSFIMIADQQALTDNARNPEKIR 60

Query: 61  ESIGNVALDYLAVGLDPNKSTLFIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK 120
            S+  VALDYLAVG+DP KST+ +QSQIPEL EL+M+Y+NLV+L+RLERNPTVK EI QK
Sbjct: 61  NSLIEVALDYLAVGIDPLKSTILVQSQIPELNELTMHYLNLVTLSRLERNPTVKAEIKQK 120

Query: 121 GFGESIPAGFLVYPVAQAADITAFKANLVPVGTDQKPMIEQTREIVRSFNHAYNCQVLVE 180
            F  SIPAGFL+YPV+QAADITAFKA  VPVG DQ PMIEQ REIVRSFN  Y  +VLVE
Sbjct: 121 NFENSIPAGFLIYPVSQAADITAFKATTVPVGEDQLPMIEQAREIVRSFNTIYGKEVLVE 180

Query: 181 PEGIYPENDAAGRLPGLDGNAKMSKSLNNGIFLADDMDTVKKKVMSMYTDPNHIKVEEPG 240
           P+ + P+    GRLPG DG AKMSKS+ N I+LAD+ D +K+KVMSMYTDPNHIKV +PG
Sbjct: 181 PKAVIPKG-TIGRLPGTDGKAKMSKSIGNAIYLADEADVIKQKVMSMYTDPNHIKVTDPG 239

Query: 241 QIEGNMVFHYLDVFGRDEDQKEITAMKEHYQKGGLGDVKTKRYLLDILERELSPIRERRL 300
           Q+EGN VF YLD F +D +  E   MK HY +GGLGDVK K++L +IL+ EL PIR RR
Sbjct: 240 QVEGNTVFTYLDTFCKDTETLE--EMKAHYSRGGLGDVKVKKFLNEILQAELEPIRNRRK 297

Query: 301 EYAKDMGQVYQMLQKGSEKAQAVAASTLDEVKSAMGLNYF 340
           E+ KD+ +VY++L++GSEKA+ VAA TL EV+  +G+ YF
Sbjct: 298 EFQKDIPEVYRILKEGSEKAREVAAGTLKEVRETIGIEYF 337
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3105> which encodes the amino acid sequence <SEQ ID 3106>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2737(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 290/340 (85%), Positives = 316/340 (92%)

Query: 1   MTKPIILTGDRPTGKLHIGHYVGSLKNRVLLQNEGSYTLFVFLADQQALTDHAKDPQTIV 60
           MTKPIILTGDRPTGKLH+GHYVGSLKNRV LQNE  Y +FVFLADQQALTDHAK+ + I
```

```
Sbjct: 2   MTKPIILTGDRPTGKLHLGHYVGSLKNRVFLQNENKYKMFVFLADQQALTDHAKESELIQ 61

Query: 61  ESIGNVALDYLAVGLDPNKSTLFIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK 120
           ESIGNVALDYL+VGLDP +ST+FIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK
Sbjct: 62  ESIGNVALDYLSVGLDPKQSTIFIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK 121

Query: 121 GFGESIPAGFLVYPVAQAADITAFKANLVPVGTDQKPMIEQTREIVRSFNHAYNCQVLVE 180
           GFGESIP+GFLVYPV+QAADITAFKANLVPVG DQKPMIEQTREIVRSFNH Y+   LVE
Sbjct: 122 GFGESIPSGFLVYPVSQAADITAFKANLVPVGNDQKPMIEQTREIVRSFNHTYHTDCLVE 181

Query: 181 PEGIYPENDAAGRLPGLDGNAKMSKSLNNGIFLADDMDTVKKKVMSMYTDPNHIKVEEPG 240
           PEGIYPEN+ AGRLPGLDGNAKMSKSL NGI+L+DD DTV+KKVMSMYTDPNHIK+E+PG
Sbjct: 182 PEGIYPENEKAGRLPGLDGNAKMSKSLGNGIYLSDDADTVRKKVMSMYTDPNHIKIEDPG 241

Query: 241 QIEGNMVFHYLDVFGRDEDQKEITAMKEHYQKGGLGDVKTKRYLLDILERELSPIRERRL 300
           QIEGNMVFHYLD+F R EDQ +I AMKEHYQ GGLGDVKTKRYLLDILEREL+PIRERRL
Sbjct: 242 QIEGNMVFHYLDIFARKEDQADIEAMKEHYQIGGLGDVKTKRYLLDILERELAPIRERRL 301

Query: 301 EYAKDMGQVYQMLQKGSEKAQAVAASTLDEVKSAMGLNYF 340
           EYAKDMG+V++MLQ+GS+KA+ VAA TL EVKSAMG+NYF
Sbjct: 302 EYAKDMGEVFRMLQEGSQKARTVAAKTLSEVKSAMGINYF 341
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1011

A DNA sequence (GBSx1079) was identified in *S.agalactiae* <SEQ ID 3107> which encodes the amino acid sequence <SEQ ID 3108>. This protein is predicted to be carbamate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0013(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA04684 GB:AJ001330 carbamate kinase [Lactobacillus sakei]
 Identities = 199/311 (63%), Positives = 254/311 (80%), Gaps = 3/311 (0%)

Query: 6   QKIVVALGGNAILSTDASAKAQQEALINTSKSLVKLIKEGHDVIVTHGNGPQVGNLLLQQ 65
           +KIVVALGGNAILSTDASA AQ +A+   T K LV   +K+G   +I++HGNGPQVGNLL+QQ
Sbjct: 4   RKIVVALGGNAILSTDASANAQIKAVKETVKQLVAFVKQGDQLIISHGNGPQVGNLLIQQ 63

Query: 66  AASDSEKNPAMPLDTCVAMTEGSIGFWLQNALNNELQEQGIDKEVATVVTQVIVDEKDQA 125
           AASDSEK PAMPLDT  AM++G IG+W+QNA N  L E+G+  +VAT+VTQ IVD KD+A
Sbjct: 64  AASDSEKTPAMPLDTVGAMSQGEIGYWMQNAFNEVLAEEGLALDVATIVTQTIVDAKDEA 123

Query: 126 FTNPTKPIGPFLSEEDAKKQAQ-ETGSKFKEDAGRGWRKVVPSPKPVGIKEASVIRRLVD 184
           F NPTKPIGPF SE +AKKQ      + F EDAGRGWR+VVPSP+P+GI+EA VI++LV+
Sbjct: 124 FQNPTKPIGPFYSEAEAKKQQSINPEAHFVEDAGRGWRRVVPSPRPIGIQEAPVIQKLVE 183

Query: 185 SGVVVISAGGGGVPVIEDANTKALKGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVFV 244
             V+ ISAGGGGVPV ++ N   L+GVEAVIDKDFAS+ L+ELV AD+ I+LT VDNV+V
Sbjct: 184 GNVITISAGGGGVPVAKEGN--KLRGVEAVIDKDFASEKLAELVGADMLIILTAVDNVYV 241

Query: 245 NFNKPNQEKLEEVTVSQMKQYITENQFAPGSMLPKVEAAIAFVENKPESRAIITSLENID 304
           NFNKP+Q+KL  V+V++++ YI ++QFA GSMLPK++ AI +V N+P+S+AIITSL+N+
Sbjct: 242 NFNKPDQKKLTNVSVAELEDYIKDDQFAKGSMLPKIQTAIEYVNNRPDSKAIITSLDNVK 301

Query: 305 NVLAQNAGTQI 315
```

```
           N+LA +AGT I
Sbjct: 302 NLLAHDAGTII 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3109> which encodes the amino acid sequence <SEQ ID 3110>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0013(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 275/312 (88%), Positives = 295/312 (94%)

Query: 6   QKIVVALGGNAILSTDASAKAQQEALINTSKSLVKLIKEGHDVIVTHGNGPQVGNLLLQQ 65
           QKIVVALGGNAILSTDASAKAQQEALI+TSKSLVKLIKEGH+VIVTHGNGPQVGNLLLQQ
Sbjct: 4   QKIVVALGGNAILSTDASAKAQQEALISTSKSLVKLIKEGHEVIVTHGNGPQVGNLLLQQ 63

Query: 66  AASDSEKNPAMPLDTCVAMTEGSIGFWLQNALNNELQEQGIDKEVATVVTQVIVDEKDQA 125
           AA+DSEKNPAMPLDTCVAMTEGSIGFWL NAL+NELQ QGI KEVA VVTQVIVD KD A
Sbjct: 64  AAADSEKNPAMPLDTCVAMTEGSIGFWLVNALDNELQAQGIQKEVAAVVTQVIVDAKDPA 123

Query: 126 FTNPTKPIGPFLSEEDAKKQAQETGSKFKEDAGRGWRKVVPSPKPVGIKEASVIRRLVDS 185
           F NPTKPIGPFL+EEDAKKQ  E+G+ FKEDAGRGWRKVVPSPKPVGIKEA+VIR LVDS
Sbjct: 124 FENPTKPIGPFLTEEDAKKQMAESGASFKEDAGRGWRKVVPSPKPVGIKEANVIRSLVDS 183

Query: 186 GVVVISAGGGGVPVIEDANTKALKGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVFVN 245
           GVVV+SAGGGGVPV+EDA +K L GVEAVIDKDFASQTLSELVDADLFIVLTGVDNV+VN
Sbjct: 184 GVVVVSAGGGGVPVVEDATSKTLTGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVYVN 243

Query: 246 FNKPNQEKLEEVTVSQMKQYITENQFAPGSMLPKVEAAIAFVENKPESRAIITSLENIDN 305
           FNKP+Q KLEEVTVSQMK+YIT++QFAPGSMLPKVEAAIAFVENKP ++AIITSLENIDN
Sbjct: 244 FNKPDQAKLEEVTVSQMKEYITQDQFAPGSMLPKVEAAIAFVENKPNAKAIITSLENIDN 303

Query: 306 VLAQNAGTQIVA 317
           VL+ NAGTQI+A
Sbjct: 304 VLSANAGTQIIA 315
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1012

A DNA sequence (GBSx1080) was identified in *S.agalactiae* <SEQ ID 3111> which encodes the amino acid sequence <SEQ ID 3112>. This protein is predicted to be permease (potE). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.63    Transmembrane  450 - 466 ( 441 - 478)
   INTEGRAL    Likelihood = -8.97    Transmembrane  236 - 252 ( 231 - 259)
   INTEGRAL    Likelihood = -8.70    Transmembrane  283 - 299 ( 277 - 308)
   INTEGRAL    Likelihood = -8.44    Transmembrane  165 - 181 ( 153 - 186)
   INTEGRAL    Likelihood = -7.96    Transmembrane  129 - 145 ( 126 - 151)
   INTEGRAL    Likelihood = -6.16    Transmembrane  396 - 412 ( 394 - 415)
   INTEGRAL    Likelihood = -5.15    Transmembrane   45 -  61 (  38 -  63)
   INTEGRAL    Likelihood = -4.94    Transmembrane  335 - 351 ( 334 - 352)
   INTEGRAL    Likelihood = -3.72    Transmembrane   13 -  29 (  10 -  30)
   INTEGRAL    Likelihood = -2.92    Transmembrane  417 - 433 ( 417 - 435)
   INTEGRAL    Likelihood = -1.54    Transmembrane  360 - 376 ( 360 - 376)
```

```
   INTEGRAL    Likelihood = -0.53   Transmembrane  207 - 223 ( 207 - 223)

----- Final Results -----
               bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10295> which encodes amino acid sequence <SEQ ID 10296> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA76779 GB:Y17554 permease [Bacillus licheniformis]
 Identities = 265/470 (56%), Positives = 347/470 (73%), Gaps = 3/470 (0%)

Query: 5   MEKEKKLGLLPLTMLVIGSLIGGGIFDLMQNMSSRAGLVPMLIAWVITAIGMGTFVLSFQ 64
            M +EKKLGL  L  LVIGS+IGGG F+L  +M+S AG   +LI W+IT +GM     SFQ
Sbjct: 1   MAEEKKLGLFALIALVIGSMIGGGAFNLASDMASGAGAGAILIGWIITGVGMIALAFSFQ 60

Query: 65  NLSEKRPDLTAGIFSYAKEGFGNFMGFNSAWGYWLSAWLGNVAYAALLFSSLGYFFKFFG 124
            NL+ KRPDL  GIF+YA+EGFG+FMGFNS WGYW +A LGNVAY  LLFS++GYF   FG
Sbjct: 61  NLTTKRPDLDGGIFTYAREGFGHFMGFNSGWGYWFAALLGNVAYGTLLFSAIGYFIPAFG 120

Query: 125 NGNNIISIIGASIVIWVVHFLILRGVNTAAFINTIVTFAKLVPVIIFLISALLAFKFNIF 184
            +G NI SIIGAS+++W VHFLILRGV +AA IN I T +KLVP+  F+I+ + F  ++F
Sbjct: 121 DGQNIASIIGASVILWCVHFLILRGVQSAAMINLITTISKLVPIFAFIIAIIFVFHLDLF 180

Query: 185 SLDIWGNGLH-QSIFNQVNSTMKTAVWVFIGIEGAVVFSGRAKKHSDIGKASILALFTMI 243
            + D WG GL   SI  QV STM   VWVF GIEGAV+FS RAKK SD+GKA+++ L +++
Sbjct: 181 TNDFWGKGLSLGSIGTQVKSTMLVTVWVFTGIEGAVLFSSRAKKSSDVGKATVIGLISVL 240

Query: 244 SLYVLISVLSLGIMSRPELANLKTPAMAYVLEKAVGHWGAILVNLGVIISVFGAILAWTL 303
             +YV+I++LSLG+M++  LA L  P+MA ++E  VG WGA+L+NLG+IISV GA LAWTL
Sbjct: 241 VIYVMITMLSLGVMNQQNLAELPNPSMAAIMEHIVGKWGAVLINLGLIISVLGAWLAWTL 300

Query: 304 FAAELPYQAAKEGAFPKFFAKENKNKAPINSLLVTNLCVQAFLITFLFTQSAYRFGFALA 363
            FA ELP  AA+EG FPK+F KENKN AP N+L +TN  +Q FL+TFL + +AY+F F+LA
Sbjct: 301 FAGELPLIAAREGVFPKWFGKENKNGAPTNALTLTNAIIQLFLLTFLISDAAYQFAFSLA 360

Query: 364 SSAILIPYAFTALYQLQFTLREDKSTPGHQKNLIIGILATIYAVYLIYAGGFDYLLLTMI 423
            SSAILIPY F+ LYQL+++    +  P   KNLIIGI+A+IY V+L+YA G DYLLLTMI
Sbjct: 361 SSAILIPYLFSGLYQLKYSWLHKE--PNRGKNLIIGIIASIYGVWLVYAAGLDYLLLTMI 418

Query: 424 AYTLGMILYIKMRKDDKLPIFVGYEKISAIVILALCLLCIIEIMTGQIDI 473
             Y  G++++  +RK  + P+F   E + A +IL L ++ +I + +G I I
Sbjct: 419 LYAPGILVFRAVRKGKEGPVFNKAELLIAALILVLAVIAVIRLASGSISI 468
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3113> which encodes the amino acid sequence <SEQ ID 3114>. Analysis of this protein sequence reveals the following:

```
      Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.52   Transmembrane  331 - 347 ( 327 - 354)
   INTEGRAL    Likelihood = -9.50   Transmembrane  390 - 406 ( 383 - 410)
   INTEGRAL    Likelihood = -8.12   Transmembrane   50 -  66 (  45 -  75)
   INTEGRAL    Likelihood = -7.59   Transmembrane  235 - 251 ( 234 - 262)
   INTEGRAL    Likelihood = -6.21   Transmembrane  133 - 149 ( 128 - 151)
   INTEGRAL    Likelihood = -5.84   Transmembrane  162 - 178 ( 153 - 183)
   INTEGRAL    Likelihood = -2.02   Transmembrane  105 - 121 ( 105 - 121)
   INTEGRAL    Likelihood = -1.49   Transmembrane  414 - 430 ( 414 - 431)
   INTEGRAL    Likelihood = -0.69   Transmembrane  280 - 296 ( 280 - 296)
   INTEGRAL    Likelihood = -0.59   Transmembrane   21 -  37 (  21 -  37)
   INTEGRAL    Likelihood = -0.32   Transmembrane  205 - 221 ( 205 - 222)

----- Final Results -----
               bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
```

```
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB85052 GB:AE000837 cationic amino acid transporter related
           protein [Methanobacterium thermoautotrophicum]
 Identities = 108/422 (25%), Positives = 213/422 (49%), Gaps = 36/422 (8%)

Query: 26  INAVIGSGIFLLPRAIYKGLGPASIAVMFGTAILTIMLAVCFAEVSGYFGKNGGAFQYSK 85
           +  ++G+ I+++       LGPASI       ++ +++A+ F+E S    + GG + Y+
Sbjct: 19  VGTIVGADIYIVAAYGAGSLGPASILAWLLAGLMALIIALVFSEASAMLPRTGGPYVYAG 78

Query: 86  RAFGDFIGFNVGFLGWTVTIFAWAAMAAGFARMFIITFPAFEGWHIPL--------SIGL 137
            A G F GF     GW++ + +W A+A     +F + F  +  + IPL         +
Sbjct: 79  EALGRFTGF---ITGWSLWVSSWVAIA-----VFPLAFIYYLEYFIPLDPPAEAVIKVLF 130

Query: 138 IILLSLMNIAGLKTSKIVTITATIAKLIPIVAFCACTLFFIKNG----LPNFTPFVQLEP 193
           I+ L+++NIAG+  +  V    TI K+ P++ F       +       + N+TP   +
Sbjct: 131 ILSLTIINIAGVGRAGKVNDILTILKVAPVLLFAVLGAIHLALNPGLLVSNYTPAAPMG- 189

Query: 194 GTNLLGAISNTAVYIFYGFIGFETLSIVAGEMRDPEKNVPRALLGSISIVSVLYMLIIGG 253
               LGA+    V +F+ ++GFE +++ A E+RDPE+ +P ++   +  V++ Y+L
Sbjct: 190 ----LGALGTVTVLVFWAYVGFELVTVPADEVRDPERTIPLSITLGMIFVTLFYILTNAV 245

Query: 254 TIAMLGSQIMMTN-APVQDAFVKMIGPAGAWMVSIGALISITGLNMGESIMVPRYGAAIA 312
            + ++  +++ ++ AP+  A   ++G  GA +++ GA+ SI G      +   R   A++
Sbjct: 246 ILGLVPWRVLASSTAPLTVAGYSLMGGIGALILTAGAVFSIAGSEEAGMLTTARLLFAMS 305

Query: 313 DEGLLPAAIAKQNQN-GAPLVAILVSGAIAIVLLLTGSFESLAKLSVVFRFFQYIPTALA 371
           ++G LP  +++ ++  G P ++ILV    A++  LTG+   L +LSVV     Y  T ++
Sbjct: 306 EDGFLPGFLSRVHRRFGTPHMSILVQNLTALLAALTGTVSGLIELSVVTLLLPYAVTCIS 365

Query: 372 VMKLRKDDPDANVIFRVPFGPIIPILAVIVSLVMIWGDNPMNFVYGAVGVIIASSVYYLM 431
           +  LR+ D            P+  +L V+V + ++    P    +G + +I++ +  YL+
Sbjct: 366 LAILRRRDGSGI--------PLKSVLGVLVCIYLLMNTTPSTTAWGLL-LILSGAPLYLI 416

Query: 432 HG 433
            G
Sbjct: 417 FG 418
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 104/368 (28%), Positives = 162/368 (43%), Gaps = 32/368 (8%)

Query: 1   MRYKMEKEKKLGLLPLTMLVIGSLIGGGIFDLMQNMSSRAGLVPMLIAWVI-TAIGMGTF 59
           M  +  ++ K  L   T+  I ++IG GIF L + +    GL P  IA +  TAI
Sbjct: 6   MNEQEREQAKFSLSGATLYGINAVIGSGIFLLPRAIYK--GLGPASIAVMFGTAILTIML 63

Query: 60  VLSFQNLSEKRPDLTAGIFSYAKEGFGNFMGFNSA---WGYWLSAWLGNVAYAALLFSSL 116
            + F  +S        G F Y+K  FG+F+GFN     W   + AW    A  A +F
Sbjct: 64  AVCFAEVSGYFGK-NGGAFQYSKRAFGDFIGFNVGFLGWTVTIFAWAAMAAGFARMFIIT 122

Query: 117 GYFFKFFGNGNNIISIIGASIVIWVVHFLILRGVNTAAFINTIVTFAKLVPVIIFLISAL 176
              F+    G +I   IG  I++ +++   + G+ T+  +    T AKL+P++ F    L
Sbjct: 123 FPAFE----GWHIPLSIGLIILLSLMN---IAGLKTSKIVTITATIAKLIPIVAFCACTL 175

Query: 177 LAFK-----FNIFSLDIWGNGLHQSIFNQVNSTMKTAVWVFIGIEGAVVFSGRAKKHSDI 231
              K     F  F     G  L  +I N       TAV++F G  G    S  A + D
Sbjct: 176 FFIKNGLPNFTPFVQLEPGTNLLGAISN-------TAVYIFYGFIGFETLSIVAGEMRDP 228

Query: 232 GKASILALFTMISLYVLISVLSLG---IMSRPELANLKTPAM-AYVLEKAVGHWGAILVN 287
            K    AL   IS+  ++ +L +G    M   ++     P   A+V  K +G  GA +V+
Sbjct: 229 EKNVPRALLGSISIVSVLYMLIIGGTIAMLGSQIMMTNAPVQDAFV--KMIGPAGAWMVS 286

Query: 288 LGVIISVFGAILAWTLFAAELPYQAAKEGAFPKFFAKENKNKAPINSLLVTNLCVQAFLI 347
           +G +IS+ G  +  ++         A EG  P   AK+N+N AP+ ++LV+       L+
Sbjct: 287 IGALISITGLNMGESIMVPRYGAAIADEGLLPAAIAKQNQNGAPLVAILVSGAIAIVLLL 346
```

```
Query: 348 TFLFTQSA 355
           T  F   A
Sbjct: 347 TGSFESLA 354
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9079> which encodes the amino acid sequence <SEQ ID 9080>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.92   Transmembrane   77 -  93 (  72 - 100)
   INTEGRAL    Likelihood = -9.29   Transmembrane  279 - 295 ( 274 - 303)
   INTEGRAL    Likelihood = -9.08   Transmembrane  203 - 219 ( 199 - 225)
   INTEGRAL    Likelihood = -8.55   Transmembrane  174 - 190 ( 171 - 197)
   INTEGRAL    Likelihood = -8.33   Transmembrane  436 - 452 ( 432 - 455)
   INTEGRAL    Likelihood = -7.32   Transmembrane  329 - 345 ( 324 - 350)
   INTEGRAL    Likelihood = -5.63   Transmembrane  402 - 418 ( 396 - 421)
   INTEGRAL    Likelihood = -4.88   Transmembrane  460 - 476 ( 456 - 479)
   INTEGRAL    Likelihood = -4.51   Transmembrane  379 - 395 ( 377 - 401)
   INTEGRAL    Likelihood = -2.81   Transmembrane   48 -  64 (  48 -  65)
   INTEGRAL    Likelihood = -2.23   Transmembrane  243 - 259 ( 243 - 262)
   INTEGRAL    Likelihood = -0.37   Transmembrane  123 - 139 ( 123 - 139)

----- Final Results -----
            bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score = 62.1 bits (148), Expect = 2e-11
 Identities = 59/250 (23%), Positives = 107/250 (42%), Gaps = 12/250 (4%)

Query: 143 WGSYLKGLLAN--YNIVLPNALNGTFNL--KNGTYIDILPV-LVMFFVTGIVLMNSKLAL 197
           WG +L   L N  Y  +L ++L   F         I I+   +V++ V  ++L     A
Sbjct: 95  WGYWLSAWLGNVAYAALLFSSLGYFFKFFGNGNNIISIIGASIVIWVVHFLILRGVNTAA 154

Query: 198 RFNSFLVILKFSALALFIFVGIFFIDHNNWSHFAPYGVGQITGGKTGIFAGASVMFFAFL 257
             N+ +   K   + +F+   +     N +S    +G G        + +      + F+
Sbjct: 155 FINTIVTFAKLVPVIIFLISALLAFKFNIFS-LDIWGNGLHQSIFNQVNSTMKTAVWVFI 213

Query: 258 GFESISMAVDEVKEPQKTIPKGIILSLIIVTALYIVVTTILTGIV---HYTKLNVPDAVA 314
           G E   +     K+    I K  IL+L  + +LY++++ +  GI+       L  P A+A
Sbjct: 214 GIEGAVVFSGRAKK-HSDIGKASILALFTMISLYVLISVLSLGIMSRPELANLKTP-AMA 271

Query: 315 FALRNIRLYWAADYVSIVAILTLITVCISMTYALARTIYSISRDGLLPKSLYTLTKKNKV 374
           + L     +W A  V++  I+++    ++ T   A   Y  +++G  PK  +    KNK
Sbjct: 272 YVLEKAVGHWGAILVNLGVIISVFGAILAWTLFAAELPYQAAKEGAFPK-FFAKENKNKA 330

Query: 375 PQNATLVTGL 384
           P N+ LVT L
Sbjct: 331 PINSLLVTNL 340
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1013

A DNA sequence (GBSx1081) was identified in *S.agalactiae* <SEQ ID 3115> which encodes the amino acid sequence <SEQ ID 3116>. This protein is predicted to be unnamed protein product (argF). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3757(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3117> which encodes the amino acid sequence <SEQ ID 3118>. Analysis of this protein sequence reveals the following:

```
        Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.48   Transmembrane  171 - 187 ( 171 - 188)

----- Final Results -----
              bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12563 GB:Z99108 similar to metabolite transporter [Bacillus subtilis]
 Identities = 190/467 (40%), Positives = 284/467 (60%), Gaps = 13/467 (2%)

Query: 25  TIFRKK-----KKYSNKTEMQRHFKVIDLVFLGLGSMVGTGIFTVTGIGAAKYAGPALTI 79
            ++FRKK        S    + R     DL  LG+G ++GTGIF +TG  AA  AGPAL I
Sbjct: 3   SLFRKKPLETLSAQSKSKSLARTLSAFDLTLLGIGCVIGTGIFVITGTVAATGAGPALII 62

Query: 80  SIIISAIAIGILALFYAEFASRMPSNGGAYSYVYATLGEFPAWLVGWYIIMEFLTAISSV 139
            S I++ +A  + A  YAEF+S +P +G  YSY Y TLGE  A+L+GW +++E++ A+S+V
Sbjct: 63  SFILAGLACALAAFCYAEFSSSIPISGSVYSYSYVTLGELLAFLIGWDLMLEYVIALSAV 122

Query: 140 AVGWGSYLKGLLANYNIVLPNALNGTFNLKNGTYIDILPVLVMFFVTGIVLMNSKLALRF 199
            A GW SY + LLA +N+ +P AL G      G   ++   +++  +T IV    K + RF
Sbjct: 123 ATGWSSYFQSLLAGFNLHIPAALTGAPGSMAGAVFNLPAAVIILLITAIVSRGVKESTRF 182

Query: 200 NSFLVILKFSALALFIFVGIFFIDHNNWSHFAPYGVGQITGGKTGIFAGASVMFFAFLGF 259
            N+ +V++K + + LFI VGI ++  +NWS F P+G+        G+   A+ +FFA+LGF
Sbjct: 183 NNVIVLMKIAIILLFIIVGIGYVKPDNWSPFMPFGM-------KGVILSAATVFFAYLGF 235

Query: 260 ESISMAVDEVKEPQKTIPKGIILSLIIVTALYIVVTTILTGIVHYTKLNVPDAVAFALRN 319
            +++S A +EVK PQK +P GII +L + T LYI V+ +LTG++ Y KLNV D V+FAL+
Sbjct: 236 DAVSNASEEVKNPQKNMPVGIISALAVCTVLYIAVSLVLTGMMPYAKLNVGDPVSFALKF 295

Query: 320 IRLYWAADYVSIVAILTLITVCISMTYALARTIYSISRDGLLPKSLYTLTKKNKVPQNAT 379
            +     A  +S+ AI+ + TV +++ YA  R  +++SRDGLLP     +    K P   T
Sbjct: 296 VGQDAVAGIISVGAIIGITTVMLALLYAQVRLTFAMSRDGLLPGLFAKVHPSFKTPFRNT 355

Query: 380 LVTGLLAMICAGIFPLSSLAEFVNICTLAYLIILSGAIIKLRRIEGEPKANEFKTPLVPF 439
             +TG++A   AG   L +LA  VN+ TLA   ++S A+I LR+   E KA+ F+ P VP
Sbjct: 356 WLTGIVAAGIAGFINLGTLAHLVNMGTLAAFTVISIAVIVLRKKHPEIKAS-FRVPFVPV 414

Query: 440 LPMLAIIICLSFMSQYKAFTWIAFAIATIIGTLIYLAYGYTHSIENK 486
            +P+++  ICL FM      TW++F I   +GTL+Y  Y   HS+ NK
Sbjct: 415 VPIISAGICLWFMYSLPGVTWLSFVIWIAVGTLVYFLYSRKHSLLNK 461
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 312/337 (92%), Positives = 324/337 (95%)

Query: 1   MTQVFQGRSFLAEKDFSREEFEYLIDFSAHLKDLKKRGVPHHYLEGKNIALLFEKTSTRT 60
            MTQVFQGRSFLAEKDF+R E EYLIDFSAHLKDLKKRGVPHHYLEGKNIALLFEKTSTRT
Sbjct: 1   MTQVFQGRSFLAEKDFTRAELEYLIDFSAHLKDLKKRGVPHHYLEGKNIALLFEKTSTRT 60

Query: 61  RAAFTTAAIDLGAHPEYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE 120
            RAAFTTAAIDLGAHPEYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE
Sbjct: 61  RAAFTTAAIDLGAHPEYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE 120
```

```
Query: 121 FSGVPVWNGLTDEWHPTQMLADYLTIKENFGKLEGITLVYCGDGRNNVANSLLVAGTLMG 180
           FSGVPVWNGLTDEWHPTQMLADY T+KENFGKLEG+TLVYCGDGRNNVANSLLV G ++G
Sbjct: 121 FSGVPVWNGLTDEWHPTQMLADYFTVKENFGKLEGLTLVYCGDGRNNVANSLLVTGAILG 180

Query: 181 VNVHIFSPKELFPAEEIVKLAEEYAKESGAHVLVTDNVDEAVKGADVFYTDVWVSMGEED 240
           VNVHIFSPKELFP EEIV LAE YAKESGA +L+T++ DEAVKGADV YTDVWVSMGEED
Sbjct: 181 VNVHIFSPKELFPEEEIVTLAEGYAKESGARILITEDADEAVKGADVLYTDVWVSMGEED 240

Query: 241 KFKERVELLQPYQVNMELIKKANNDNLIFLHCLPAFHDTNTVYGKDVAEKFGVKEMEVTD 300
           KFKERVELLQPYQVNM+L++KA ND LIFLHCLPAFHDTNTVYGKDVAEKFGVKEMEVTD
Sbjct: 241 KFKERVELLQPYQVNMDLVQKAGNDKLIFLHCLPAFHDTNTVYGKDVAEKFGVKEMEVTD 300

Query: 301 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV 337
           EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV
Sbjct: 301 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV 337
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1014

A DNA sequence (GBSx1082) was identified in *S.agalactiae* <SEQ ID 3119> which encodes the amino acid sequence <SEQ ID 3120>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0456(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10921> which encodes amino acid sequence <SEQ ID 10922> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3121> which encodes the amino acid sequence <SEQ ID 3122>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.41   Transmembrane  121 - 137 ( 118 - 140)

----- Final Results -----
               bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 65/113 (57%), Positives = 83/113 (72%)

Query: 31  MEEEFDDNDEQDTIYAVLYDGKQPVSTGRFLPETQTEARLTRIATLKGYRGNGYGTKIII 90
           M ++FD NDE  T+YAV+YD  QPVSTG+FL ET+ EARLTRI TL  Y G GYG K+
Sbjct: 1   MADKFDANDETRTVYAVVYDNDQPVSTGQFLAETKIEARLTRIVTLADYCGCGYGAKVTE 60

Query: 91  ALENYAKENGYHYLTIHAELTAKDFYQTLGYQATGNIYMEDGEACQTLEKYLI 143
           ALE Y +  G++ LTIH+ELTA+ FY+ LGYQ+ G   +EDGE CQ+L K ++
Sbjct: 61  ALETYTRREGFYQLTIHSELTAQTFYENLGYQSYGPKCLEDGEYCQSLAKTIL 113
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1015

A DNA sequence (GBSx1083) was identified in *S.agalactiae* <SEQ ID 3123> which encodes the amino acid sequence <SEQ ID 3124>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3125> which encodes the amino acid sequence <SEQ ID 3126>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2730(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 375/411 (91%), Positives = 395/411 (95%), Gaps = 1/411 (0%)

Query: 1   MTQTHPIHVFSEIGKLKKVMLHRPGKEIENLMPDYLERLLFDDIPFLEDAQKEHDAFAQA 60
           MT   PIHV+SEIGKLKKV+LHRPGKEIENLMPDYLERLLFDDIPFLEDAQKEHDAFAQA
Sbjct: 1   MTAQTPIHVYSEIGKLKKVLLHRPGKEIENLMPDYLERLLFDDIPFLEDAQKEHDAFAQA 60

Query: 61  LRNEGVEVLYLENLAAESLTNQEIREQFIDEYIGEANVRGRATKKAIRELLLNIKDNKEL 120
           LR+EG+EVLYLE LAAESL   EIRE FIDEY+ EAN+RGRATKKAIRELL+ I+DN+EL
Sbjct: 61  LRDEGIEVLYLETLAAESLVTPEIREAFIDEYLSEANIRGRATKKAIRELLMAIEDNQEL 120

Query: 121 IEKTMAGIQKSELPEIPSSEKGLTDLVESNYPFAIDPMPNLYFTRDPFATIGNGVSLNHM 180
           IEKTMAG+QKSELPEIP+SEKGLTDLVESNYPFAIDPMPNLYFTRDPFATIG GVSLNHM
Sbjct: 121 IEKTMAGVQKSELPEIPASEKGLTDLVESNYPFAIDPMPNLYFTRDPFATIGTGVSLNHM 180

Query: 181 FSETRNRETLYGKYIFTHHPEYGG-KVPMVYEREETTRIEGGDELVLSKDVLAVGISQRT 239
           FSETRNRETLYGKYIFTHHP YGG KVPMVY+R ETTRIEGGDELVLSKDVLAVGISQRT
Sbjct: 181 FSETRNRETLYGKYIFTHHPIYGGGKVPMVYDRNETTRIEGGDELVLSKDVLAVGISQRT 240

Query: 240 DAASIEKLLVNIFKQNLGFKKVLAFEFANNRKFMHLDTVFTMVDYDKFTIHPEIEGDLRV 299
           DAASIEKLLVNIFKQNLGFKKVLAFEFANNRKFMHLDTVFTMVDYDKFTIHPEIEGDLRV
Sbjct: 241 DAASIEKLLVNIFKQNLGFKKVLAFEFANNRKFMHLDTVFTMVDYDKFTIHPEIEGDLRV 300

Query: 300 YSVTYENQDLHIEEEKGDLADLLAKNLGVEKVELIRCGGDNLVAAGREQWNDGSNTLTIA 359
           YSVTY+N++LHI EEKGDLA+LLA NLGVEKV+LIRCGGDNLVAAGREQWNDGSNTLTIA
Sbjct: 301 YSVTYDNEELHIVEEKGDLAELLAANLGVEKVDLIRCGGDNLVAAGREQWNDGSNTLTIA 360

Query: 360 PGVVIVYNRNTITNAILESKGLKLIKINGSELVRGRGGPRCMSMPFEREDL 410
           PGVV+VYNRNTITNAILESKGLKLIKI+GSELVRGRGGPRCMSMPFERED+
Sbjct: 361 PGVVVVYNRNTITNAILESKGLKLIKIHGSELVRGRGGPRCMSMPFEREDI 411
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1016

A DNA sequence (GBSx1084) was identified in *S.agalactiae* <SEQ ID 3127> which encodes the amino acid sequence <SEQ ID 3128>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3162(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8703> which encodes amino acid sequence <SEQ ID 8704> was also identified. This protein has an RGD motif and has homology with the following sequences in the GENPEPT database.

```
>GP:AAG07568 GB:AE004834 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 42/132 (31%), Positives = 74/132 (55%), Gaps = 3/132 (2%)

Query: 35  IQTYRKAYQTFKTK-KGARSSIEALLKRVNSGNEITSINPLVDIYNAASLRFGLPIGAED 93
            +  + +A++ F  K +    S EAL KR      + SI+P+VD+YNA S++F +P+G E+
Sbjct: 63  LAAWAEAFRRFGAKPQRTPCSAEALRKRALRDGGLPSIDPVVDLYNAISVQFAIPVGGEN 122

Query: 94  SDTFRGDLKLTITNGGDEFYLI--GEDFNRPTLSGELAYVDDVGAVCRCFNWRDGKRTMI 151
               + G  +L + +G + F  +  GE  +     GE+ + DD+G  CR +NWR G RT +
Sbjct: 123 LAAYAGPPRLVVADGSETFDTLKNGEALDESPDPGEVVWRDDLGVTCRRWNWRQGVRTRL 182

Query: 152 TDNTQNAFLVIE 163
              + +  + ++E
Sbjct: 183 DASARRMWFILE 194
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3129> which encodes the amino acid sequence <SEQ ID 3130>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0700(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 127/199 (63%), Positives = 155/199 (77%)

Query: 8   ELKQLLSDSHSLAKKYLQEKEFSQNRVIQTYRKAYQTFKTKKGARSSIEALLKRVNSGNE 67
           ++KQLL+DSH LAK YL    FS N+V+Q YRKAYQ FKTKKGARSSIEALLKRV++G
Sbjct: 36  DVKQLLADSHELAKAYLTADNFSDNQVVQVYRKAYQHFKTKKGARSSIEALLKRVSNGQS 95

Query: 68  ITSINPLVDIYNAASLRFGLPIGAEDSDTFRGDLKLTITNGGDEFYLIGEDFNRPTLSGE 127
           I SINPLVDIYNAASLRFGLP GAEDSD+F GDL+LTIT+GGD+FYLIG+  N PTL  E
Sbjct: 96  IPSINPLVDIYNAASLRFGLPAGAEDSDSFIGDLRLTITDGGDDFYLIGDADNNPTLPNE 155

Query: 128 LAYVDDVGAVCRCFNWRDGKRTMITDNTQNAFLVIELIDNGREIIFKEALDFIATNTNRF 187
           L Y DD+GA CRC NWRDG+RTM+T++T+NAFL+IE +D   +   +EAL FI  +   +
Sbjct: 156 LCYKDDIGAFCRCLNWRDGERTMVTEHTKNAFLIIEALDQEGQNRLQEALKFIEGSAKMY 215

Query: 188 LKAKTQTIILDKEHSEITL 206
           L A T   +LDK++  + L
Sbjct: 216 LHAITSVHVLDKDNPHVPL 234
```

SEQ ID 8704 (GBS298) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 2; MW 29kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 5; MW 54kDa).

The GBS298-GST fusion product was purified (Figure 203, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 297), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1017

A DNA sequence (GBSx1085) was identified in *S.agalactiae* <SEQ ID 3131> which encodes the amino acid sequence <SEQ ID 3132>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3770(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1018

A DNA sequence (GBSx1086) was identified in *S.agalactiae* <SEQ ID 3133> which encodes the amino acid sequence <SEQ ID 3134>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4263(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB95946 GB:Y17554 Crp/Fnr family protein [Bacillus
           licheniformis]
 Identities = 85/214 (39%), Positives = 126/214 (58%), Gaps = 14/214 (6%)

Query: 11  RQLDDFKHFTIEQFDHIVSHIKHRTALKNHTLFFEGDYREKLFLIQSGHVKIEQSDASGS 70
           R L+D K F           I  R+  K   LF E D RE+++L+  G +K+E+S+ +GS
Sbjct: 22  RDLEDMKQF-----------IYWRSYHKGQILFMEDDPRERMYLLLDGFIKLEKSNEAGS 70

Query: 71  FIYTDYVRQGTVFPYGGLFLDDDYHFSAVAITDIEYFSLPMALYEEYSLQNINQMKHLCR 130
             YTDYVR  T+FP+GGLF D+ YH++A A+TDIE + +PM ++E+    N N +  +
Sbjct: 71  MFYTDYVRPHTLFPFGGLFRDEHYHYAAEALTDIELYYIPMNIFEDLVRDNKNLLYDILN 130

Query: 131 KYSKLLRVHEIRLRNMVTSSASMRVIQSLATL---LLQVPTERGHLPFPITTIEIANMSG 187
             S +L +HE RL+ +  S A  RV Q++  L   L Q  +    +  PIT  EIA +SG
Sbjct: 131 HLSDILALHEERLKRITLSHAHDRVTQAIYYLTESLGQKESNSTVINCPITAAEIAKISG 190

Query: 188 TTRETVSHVLKELRQKDIVEMKGKKLLYNNKNYF 221
           T+RETVS VLK+LR + ++    K+++ N   YF
Sbjct: 191 TSRETVSAVLKKLRCEGVISQMNKQIMINRPEYF 224
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3135> which encodes the amino acid sequence <SEQ ID 3136>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4478(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 130/224 (58%), Positives = 180/224 (80%)

Query: 1   MITKEQYFYFRQLDDFKHFTIEQFDHIVSHIKHRTALKNHTLFFEGDYREKLFLIQSGHV 60
            +I +E Y Y R+L+DF++F+IEQFD IV  ++ R A K+H LFFEGD R+KLFL+ SG+
Sbjct: 1   VIRREDYQYLRKLNDFRYFSIEQFDKIVGQMEFRKAKKDHILFFEGDKRDKLFLVTSGYF 60

Query: 61  KIEQSDASGSFIYTDYVRQGTVFPYGGLFLDDDYHFSAVAITDIEYFSLPMALYEEYSLQ 120
            K+EQSD SG+F+YTD++R GT+FPYGGLF DD YHFS VA+TD+ YF  P+ L+E+YSL+
Sbjct: 61  KVEQSDQSGTFMYTDFIRHGTIFPYGGLFTDDYYHFSVVAMTDVTYFYFPVDLFEDYSLE 120

Query: 121 NINQMKHLCRKYSKLLRVHEIRLRNMVTSSASMRVIQSLATLLLQVPTERGHLPFPITTI 180
            N  QMKHL  K SKLL +HE+R+RN++TSSAS RVIQSLA LL+++  +   LPF +TT
Sbjct: 121 NRLQMKHLYSKMSKLLELHELRVRNLITSSASSRVIQSLAILLVEMGKDSDTLPFQLTTT 180

Query: 181 EIANMSGTTRETVSHVLKELRQKDIVEMKGKKLLYNNKNYFKKF 224
            +IA +SGTTRETVSHVL++L++++++ +KGK L Y +K+YF ++
Sbjct: 181 DIAQISGTTRETVSHVLRDLKKQELITIKGKYLTYLDKDYFLQY 224
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1019

A DNA sequence (GBSx1087) was identified in *S.agalactiae* <SEQ ID 3137> which encodes the amino acid sequence <SEQ ID 3138>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1643(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2161> which encodes the amino acid sequence <SEQ ID 2162>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1201(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 461/493 (93%), Positives = 478/493 (96%)

Query: 2   MSNWDTKFLKKGFTFDDVLLIPAESHVLPNEVDMKTKLADNLTLNIPIITAAMDTVTDSK 61
            MSNWDTKFLKKG+TFDDVLLIPAESHVLPNEVD+KTKLADNLTLNIPIITAAMDTVT SK
```

```
Sbjct: 1   MSNWDTKFLKKGYTFDDVLLIPAESHVLPNEVDLKTKLADNLTLNIPIITAAMDTVTGSK 60

Query: 62  MAIAIARAGGLGIIHKNMSIVDQAEEVRKVKRSENGVIIDPFFLTPDNTVSEAEELMQNY 121
           MAIAIARAGGLG+IHKNMSI +QAEEVRKVKRSENGVIIDPFFLTP++ VSEAEELMQ Y
Sbjct: 61  MAIAIARAGGLGVIHKNMSITEQAEEVRKVKRSENGVIIDPFFLTPEHKVSEAEELMQRY 120

Query: 122 RISGVPIVETLENRKLVGIITNRDMRFISDYKQLISEHMTSQNLVTAPIGTDLETAERIL 181
           RISGVPIVETL NRKLVGIITNRDMRFISDY   ISEHMTS++LVTA +GTDLETAERIL
Sbjct: 121 RISGVPIVETLANRKLVGIITNRDMRFISDYNAPISEHMTSEHLVTAAVGTDLETAERIL 180

Query: 182 HEHRIEKLPLVDDEGRLSGLITIKDIEKVIEFPKAAKDEFGRLLVAGAVGVTSDTFERAE 241
           HEHRIEKLPLVD+ GRLSGLITIKDIEKVIEFP AAKDEFGRLLVA AVGVTSDTFERAE
Sbjct: 181 HEHRIEKLPLVDNSGRLSGLITIKDIEKVIEFPHAAKDEFGRLLVAAAVGVTSDTFERAE 240

Query: 242 ALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPNRTLIAGNIATAEGARALYDAGVDVV 301
           ALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPNRTLIAGNIATAEGARALYDAGVDVV
Sbjct: 241 ALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPNRTLIAGNIATAEGARALYDAGVDVV 300

Query: 302 KVGIGPGSICTTRVVAGVGVPQITAIYDAAAVAREYGKTIIADGGIKYSGDIVKALAAGG 361
           KVGIGPGSICTTRVVAGVGVPQ+TAIYDAAAVAREYGKTIIADGGIKYSGDIVKALAAGG
Sbjct: 301 KVGIGPGSICTTRVVAGVGVPQVTAIYDAAAVAREYGKTIIADGGIKYSGDIVKALAAGG 360

Query: 362 NAVMLGSMFAGTDEAPGETEIFQGRKFKTYRGMGSIAAMKKGSSDRYFQGSVNEANKLVP 421
           NAVMLGSMFAGTDEAPGETEI+QGRKFKTYRGMGSIAAMKKGSSDRYFQGSVNEANKLVP
Sbjct: 361 NAVMLGSMFAGTDEAPGETEIYQGRKFKTYRGMGSIAAMKKGSSDRYFQGSVNEANKLVP 420

Query: 422 EGIEGRVAYKGSVADIVFQMLGGIRSGMGYVGAANIKELHDNAQFVEMSGAGLKESHPHD 481
           EGIEGRVAYKG+ +DIVFQMLGGIRSGMGYVGA +I+ELH+NAQFVEMSGAGL ESHPHD
Sbjct: 421 EGIEGRVAYKGAASDIVFQMLGGIRSGMGYVGAGDIQELHENAQFVEMSGAGLIESHPHD 480

Query: 482 VQITNEAPNYSVH 494
           VQITNEAPNYSVH
Sbjct: 481 VQITNEAPNYSVH 493
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1020

A DNA sequence (GBSx1089) was identified in *S.agalactiae* <SEQ ID 3139> which encodes the amino acid sequence <SEQ ID 3140>. This protein is predicted to be MutR. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1841(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD04237 GB:AF007761 MutR [Streptococcus mutans]
 Identities = 51/215 (23%), Positives = 102/215 (46%), Gaps = 9/215 (4%)

Query: 5   GKILKELREDKGISLSSLAKSAQLSKSTLSRFENGETQIGIDKFIKALQTLEVGVTINEV 64
           G++ KELR  +G+ L  +A+   LS S LS+FENG+T +  DK I A+Q +   +T +E
Sbjct: 9   GELYKELRMARGLKLKDIARD-NLSVSQLSKFENGQTMLAADKLILAIQGIH--MTFSEF 65

Query: 65  SILDSKVKAGTSNTDLEQLTLLESYRDNEDIMRIFSFQKQQSCDRIESNVLKILAKLFIS 124
           S   ++ +        ++L  L++ +D + + +I           + + + K++ K  +
Sbjct: 66  SYAFTQYQESDLFKTGKKLVELQTKKDIKGLKKILKDYPDTETYNVYNRLNKLVIKAAVY 125

Query: 125 NLGLNMRLPQDEINLVVTYLNGVTQYNDFYFKVICYFQDILPED--VILNKI----SNMT 178
```

```
           +L  +  +  +E   + +YL  + ++ ++   +      IL +D  V L K       +
Sbjct: 126 SLDSSFEITNEEKEFLTSYLYAIEEWTEYELYLFGNTLFILSDDDLVFLGKAFVERDKLY 185

Query: 179 KEQLPYSKSLVNLLIKQVIIALEKDSVDKAIVFAD 213
           +E   + K    +LI  ++I +E  S   A  F +
Sbjct: 186 RELSEHKKRAELVLINLILILVEHHSFYHAQYFIE 220
```

There is also homology to SEQ ID 628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1021

A DNA sequence (GBSx1090) was identified in *S.agalactiae* <SEQ ID 3141> which encodes the amino acid sequence <SEQ ID 3142>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.77   Transmembrane  269 - 285 ( 265 - 287)
   INTEGRAL    Likelihood = -6.90   Transmembrane   33 -  49 (  31 -  51)
   INTEGRAL    Likelihood = -6.79   Transmembrane  182 - 198 ( 176 - 200)
   INTEGRAL    Likelihood = -6.37   Transmembrane  117 - 133 ( 113 - 135)
   INTEGRAL    Likelihood = -5.57   Transmembrane  240 - 256 ( 232 - 259)
   INTEGRAL    Likelihood = -3.40   Transmembrane  223 - 239 ( 220 - 239)
   INTEGRAL    Likelihood = -0.96   Transmembrane   56 -  72 (  55 -  72)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3143> which encodes the amino acid sequence <SEQ ID 3144>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.99   Transmembrane  269 - 285 ( 264 - 286)
   INTEGRAL    Likelihood = -8.76   Transmembrane  117 - 133 ( 112 - 135)
   INTEGRAL    Likelihood = -7.70   Transmembrane  179 - 195 ( 174 - 200)
   INTEGRAL    Likelihood = -4.83   Transmembrane   34 -  50 (  32 -  52)
   INTEGRAL    Likelihood = -4.46   Transmembrane  213 - 229 ( 211 - 230)
   INTEGRAL    Likelihood = -4.14   Transmembrane  240 - 256 ( 232 - 259)
   INTEGRAL    Likelihood = -0.69   Transmembrane   91 - 107 (  91 - 108)
   INTEGRAL    Likelihood = -0.32   Transmembrane    4 -  20 (   4 -  20)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9181> which encodes the amino acid sequence <SEQ ID 9182>. Analysis of this protein sequence reveals the following:

```
     Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
     INTEGRAL    Likelihood =-10.99   Transmembrane  259 - 275 ( 254 - 276)
     INTEGRAL    Likelihood = -8.76   Transmembrane  107 - 123 ( 102 - 125)
     INTEGRAL    Likelihood = -7.70   Transmembrane  169 - 185 ( 164 - 190)
     INTEGRAL    Likelihood = -4.83   Transmembrane   24 -  40 (  22 -  42)
     INTEGRAL    Likelihood = -4.46   Transmembrane  203 - 219 ( 201 - 220)
     INTEGRAL    Likelihood = -4.14   Transmembrane  230 - 246 ( 222 - 249)
     INTEGRAL    Likelihood = -0.69   Transmembrane   81 -  97 (  81 -  98)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.539(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 200/287 (69%), Positives = 244/287 (84%)

Query: 1   MEGLLIALIPMFAWGSIGFVSNKIGGRPNQQTFGMTLGALLFAIIVWLFKQPEMTASLWI 60
           +EG+  ALIPMF WGSIGFVSNKIGG+P+QQT GMT GALLF++ VWL  +PEMT  LW+
Sbjct: 1   LEGIFYALIPMFTWGSIGFVSNKIGGKPSQQTLGMTFGALLFSLAVWLIVRPEMTLQLWL 60

Query: 61  FGILGGILWSVGQNGQFQAMKYMGVSVANPLSSGAQLVGGSLVGALVFHEWTKPIQFILG 120
           FGILGG +WS+GQ GQF AM+YMGVSVANPLSSG+QLV GSL+G LVFHEWT+P+QF++G
Sbjct: 61  FGILGGFIWSIGQTGQFHAMQYMGVSVANPLSSGSQLVLGSLIGVLVFHEWTRPMQFVVG 120

Query: 121 LTALTLLVIGFYFSSKRDVSEQALATHQEFSKGFATIAYSTVGYISYAVLFNNIMKFDAM 180
             AL LL++GFYFSSK+D +   +     FSKGF  + YST+GY+ YAVLFNNIMKF+ +
Sbjct: 121 SLALLLLIVGFYFSSKQDDANAQVNHLHNFSKGFRALTYSTIGYVMYAVLFNNIMKFEVL 180

Query: 181 AVILPMAVGMCLGAICFMKFRVNFEAVVVKNMITGLMWGVGNVFMLLAAAKAGLAIAFSF 240
           +VILPMAVGM LGAI FM F+++ +  V+KN + GL+WG+GN+FMLLAA+KAGLAIAFSF
Sbjct: 181 SVILPMAVGMVLGAITFMSFKISIDQYVIKNSVVGLLWGIGNIFMLLAASKAGLAIAFSF 240

Query: 241 SQLGVIISIIGGILFLGETKTKKEQKWVVMGILCFVMGAILLGIVKS 287
           SQLG IISI+GGILFLGETKTKKE +WVV GI+CF++GAILLG+VKS
Sbjct: 241 SQLGAIISIVGGILFLGETKTKKEMRWVVTGIICFIVGAILLGVVKS 287
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1022

A DNA sequence (GBSx1092) was identified in *S.agalactiae* <SEQ ID 3145> which encodes the amino acid sequence <SEQ ID 3146>. This protein is predicted to be recf protein (recF). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2653(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3147> which encodes the amino acid sequence <SEQ ID 3148>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1677(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 248/364 (68%), Positives = 300/364 (82%), Gaps = 1/364 (0%)

Query: 1   MWIKNISLKHYRNYEEAQVDFSPNLNIFIGRNAQGKTNFLEAIYFLALTRSHRTRSDKEL 60
           MWIK + LKHYRNY+     FS  LN+FIG NAQGKTNFLEAIYFL+LTRSHRTR+DKEL
Sbjct: 1   MWIKELELKHYRNYDHLLASFSSGLNVFIGNNAQGKTNFLEAIYFLSLTRSHRTRADKEL 60
```

```
Query: 61  VHFKHHDVQITGEVIRKSGHLNLDIQLSEKGRITKVNHLKQAKLSDYIGAMTVVLFAPED 120
           +HF H  V +TG++ R SG ++L+I LS+KGR+TK+N LKQAKLSDYIG M VVLFAPED
Sbjct: 61  IHFDHSTVSLTGKIQRISGTVDLEINLSDKGRVTKINALKQAKLSDYIGTMMVVLFAPED 120

Query: 121 LQLVKGAPSLRRKFLDIDIGQIKPTYLAELSNYNHVLKQRNTYLKTTNNVDKTFLTVLDE 180
           LQLVKGAPSLRRKF+DID+GQIKP YL+ELS+YNHVLKQRN+YLK+   +D  FL VLDE
Sbjct: 121 LQLVKGAPSLRRKFIDIDLGQIKPVYLSELSHYNHVLKQRNSYLKSAQQIDAAFLAVLDE 180

Query: 181 QLADYGSRVIEHRFDFIQALNDEADKHHYIISTELEHLSIHYKSSIEFTDKSSIREHFLN 240
           QLA YG+RV+EHR DFI AL  EA+ HH  IS  LE LS+ Y+SS+ F  K++I + FL+
Sbjct: 181 QLASYGARVMEHRIDFINALEKEANTHHQAISNGLESLSLSYQSSVVFDKKTNIYQQFLH 240

Query: 241 QLSKSHSRDIFKKNTSIGPHRDDITFFINDINATFASQGQQRSLILSLKLAEIELIKTVT 300
           QL K+H +D F+KNTS+GPHRD++ F+IN +NA FASQGQ RSLILSLK+AE+ L+K +T
Sbjct: 241 QLEKNHQKDFFRKNTSVGPHRDELAFYINGMNANFASQGQHRSLILSLKMAEVSLMKALT 300

Query: 301 NDYPILLLDDVMSELDNHRQLKLLEG-IKENVQTFITTTSLEHLSALPDQLKIFNVSDGT 359
            D PILLLDDVMSELDN RQ KLLE  IKENVQTFITTTSL+HLS LP+ ++IF+V+ GT
Sbjct: 301 GDNPILLLDDVMSELDNTRQTKLLETVIKENVQTFITTTSLDHLSQLPEGIRIFHVTKGT 360

Query: 360 ISIN 363
           + I+
Sbjct: 361 VQID 364
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1023

A DNA sequence (GBSx1093) was identified in *S.agalactiae* <SEQ ID 3149> which encodes the amino acid sequence <SEQ ID 3150>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1807(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA61548 GB:X89367 orf121 [Lactococcus lactis]
 Identities = 56/116 (48%), Positives = 74/116 (63%), Gaps = 3/116 (2%)

Query: 3   YKLFDEYITLQSLLKEIGIIQSGGAIKKFLADNR--VLFNGDLENRRGKKLRLGDIITIP 60
           Y LF+EYITL  LLKE+G+I +GG  K FLA+N   + +NG+ ENRRGKKLR GD++  P
Sbjct: 4   YILFEEYITLGQLLKELGLISTGGQPKIFLAENEGNIFYNGEAENRRGKKLRDGDLLEFP 63

Query: 61  DQNIEIIIRKPSDQEIEERNIEIAEKQRVSAIVKEMNKNTNKGKSKTSKKPVRFPG 116
             ++++   +     I+E   E AE+ RV AIVK+MN   NK      K P RFPG
Sbjct: 64  TFDLKVTFEQADADAIKEHEAEKAEEARVKAIVKKMNAE-NKTTKPAKKAPPRFPG 118
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3151> which encodes the amino acid sequence <SEQ ID 3152>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0493(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 74/136 (54%), Positives = 94/136 (68%), Gaps = 20/136 (14%)

Query: 1   MDYKLFDEYITLQSLLKEIGIIQSGGAIKKFLADNRVLFNGDLENRRGKKLRLGDIITIP 60
           M YKLF E+ITLQ+LLKE+GIIQSGGAIK FLA+  VLFNG+ E RRGKK+R+GD I++P
Sbjct: 9   MIYKLFTEFITLQALLKELGIIQSGGAIKGFLAETTVLFNGEDEKRRGKKIRVGDKISLP 68

Query: 61  DQNIEIIIRKPSDQEIEERNIEIAEKQRVSAIVKEMNKNTNKGKSK------TSKK---- 110
           DQ++ I I +PS +E E+   E+AEK RV+A+VK+MN+   K  SK      T+KK
Sbjct: 69  DQDLIITIVEPSQEEKEQFAEEMAEKTRVAALVKQMNQANKKTSSKHNNRQSTTKKSLRA 128

Query: 111 ----------PVRFPG 116
                     PVRFPG
Sbjct: 129 TKKTKGKPTAPVRFPG 144
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1024

A DNA sequence (GBSx1094) was identified in *S.agalactiae* <SEQ ID 3153> which encodes the amino acid sequence <SEQ ID 3154>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.86   Transmembrane  269 - 285 ( 267 - 285)

----- Final Results -----
               bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3155> which encodes the amino acid sequence <SEQ ID 3156>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3008(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 227/413 (54%), Positives = 309/413 (73%)

Query: 1   MKIVEGVSLHLIKNQQFKTNHLTFRFSGDFNNKTVARRSLVAQMLVTANAKYPKVQEFRE 60
           MKIV+GV LHLIK +QFKTNH+TFRFSGD N KTVA++ LVAQML TAN  YP V++FRE
Sbjct: 1   MKIVQGVQLHLIKTKQFKTNHITFRFSGDLNQKTVAKKVLVAQMLATANECYPTVRQFRE 60

Query: 61  KLASLYGASLSTKISTKGLVHIVDIDIVFVKNTFTLEQENIVEQIITFLEDMLFSPLISL 120
           KLA LYGASLST + TKGLVHIVDIDI F+++ +    E I++++I FL+D+LFSPL+S+
Sbjct: 61  KLARLYGASLSTNVLTKGLVHIVDIDITFIQDRYACNGEKILDEMIQFLKDILFSPLLSI 120

Query: 121 EQYQTSIFDTEKKNLIQYLEADIEDNFYSSDLALKSLFYNNKTLRLPKYGTASLVESENS 180
            QYQ  +F+TEK NLI Y+E+D ED+FY S L +K LFY NK L++ +YG+  L+  E +
Sbjct: 121 AQYQPKVFETEKNNLINYIESDREDSFYYSSLKVKELFYCNKNLQMSEYGSPELIAKETA 180

Query: 181 FTAYQEFQKMLKEDQLDIFVVGDFDDYRMIQAFNRMAFEPRHKVLAFDYTQTYENITRSQ 240
           +T+YQEF KML EDQ+DIF++GDFDDYR++Q  ++   + R+K L F + Q   NI +
Sbjct: 181 YTSYQEFHKMLNEDQIDIFILGDFDDYRVVQLIHQFPLDNRNKNLNFFHLQNSVNIIKES 240

Query: 241 VEDKDVNQSIMQLAYHLPITYKDEDYFALIVFNGLFGAFAHSLLFTEIREKQGLAYTIGS 300
```

```
           +E + V+QSI+QLAYH P  +   DY+AL++ NGL G+FAHS LF +IRE++GLAY+IG
Sbjct: 241 IEKRAVHQSILQLAYHFPSVFGQRDYYALVLLNGLLGSFAHSRLFIKIREEEGLAYSIGC 300

Query: 301 QFDSFTGLFTIYAGIDKENRERFLKLINKQFNNIKMGRFSSTLLKQTKDILKMNYVLASD 360
           +FDS+TGLF IY GID ++R + L+LI ++ N IKMGRFS  L+K+T+ +L  N +L+ D
Sbjct: 301 RFDSYTGLFEIYTGIDSQHRTKTLQLIIQELNAIKMGRFSEQLIKKTRSMLLNNALLSED 360

Query: 361 NPKVIVDHIYHEHYLDQFHTSALFIDKVDDVTKSDIVSVATKLKLQAFYFLEG 413
             K I++ IY   Y+D  ++   +I  V++V K+DI+ VA  LKLQ  YFLEG
Sbjct: 361 YNKNIIERIYRSSYIDSSYSIKNWIKGVNEVNKADIIKVANLLKLQTVYFLEG 413
```

SEQ ID 3154 (GBS400) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 2; MW 49.2kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 3; MW 74kDa) and in Figure 177 (lane 6; MW 74kDa).

GBS400-GST was purified as shown in Figure 217, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1025

A DNA sequence (GBSx1095) was identified in *S.agalactiae* <SEQ ID 3157> which encodes the amino acid sequence <SEQ ID 3158>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3473(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3159> which encodes the amino acid sequence <SEQ ID 3160>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4298(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 207/424 (48%), Positives = 276/424 (64%), Gaps = 3/424 (0%)

Query: 5   KITYQNLQEEVYKLTLESGLNVYLIPKPSFKETVGVLTANFGSLHTKYTRNGCVEHYPAG 64
           KI Y N+ E++Y + LE+GL VY I K  F E   +LT  FGSL  K T +      PAG
Sbjct: 6   KINYPNIDEDLYYVKLENGLTVYFIKKIGFLEKTAMLTVGFGSLDNKLTVDDESRDAPAG 65

Query: 65  IAHFLEHKLFELDKGQDAATQFTKYGAESNAFTTFDKTSFYFSTISHITNCLDILLDFVL 124
           IAHFLEHKLFE + G D + +FT+ GAE+NAFTTF++TSF+FST S     L++L  FVL
Sbjct: 66  IAHFLEHKLFEDESGGDISLKFTQLGAETNAFTTFNQTSFFFSTASKFQENLELLQYFVL 125

Query: 125 TTNFTEESITKEKDIIKQEIEMYQDDPEYRLYQGVLSNLYPNSPLAFDIAGDYQSISQIT 184
           + N T+ES+++EK II QEI+MYQDD +YR Y G+L NL+P + LA DIAG   SI +IT
Sbjct: 126 SANITDESVSREKKIIGQEIDMYQDDADYRAYSGILQNLFPKTSLANDIAGSKASIQKIT 185

Query: 185 LTDLQENHKDFYQLSNMNLVLVGQFSPQEIITYLQKNSHFTSY--SQNIDRDSISLEPVI 242
              L+ +H  FYQ +NM+L +VG     E    +Q+     SY   + +  D +   PVI
```

```
Sbjct: 186 KILLETHHTYFYQPTNMSLFIVGDIDIDETFLAIQRFQTTLSYPDRKRVTVDPLHYYPVI 245

Query: 243 KNNSCHMTVTKPKLAIGYRKSNHMIHGSYLKEKIGLQLFFAMLLGWTSTINQDWYESGQI 302
           K++S  M VT  KL +G+R    +   S L  +I L+LF +ML+GWTS I    YE G+I
Sbjct: 246 KSSSVDMDVTTAKLVVGFRGYLTLTQHSLLTYRIALKLFLSMLIGWTSKIYHTLYEDGKI 305

Query: 303 DDSFDIEIEVHPDFECVIISLDTTEPIAFSTQLRLLLKNALQSSDLTESHLKNVKRELYG 362
           DDSFD+++E+H +F+ V+ISLDT EPIA S  +R  L     S + T  HL  +K+E+YG
Sbjct: 306 DDSFDVDVEIHHNFQFVLISLDTPEPIAMSNYIRQKLATIKISKEFTNEHLNLLKKEMYG 365

Query: 363 DFLRSLDSIENLAMQFVTYLYDG-KTMYLDLPSIVEELDLEDVITIGKDFLDNADTSDFV 421
           DF++SLDSIE+L  QF  YL D  K  Y D+P I+E L L+DV+TIGK F + AD SDF
Sbjct: 366 DFIQSLDSIEHLTHQFSLYLSDSDKETYFDIPKIIERLTLKDVVTIGKAFFEKADASDFT 425

Query: 422 IFPK 425
           +FPK
Sbjct: 426 VFPK 429
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1026

A DNA sequence (GBSx1096) was identified in *S.agalactiae* <SEQ ID 3161> which encodes the amino acid sequence <SEQ ID 3162>. This protein is predicted to be phosphotidylglycerophosphate synthase (pgsA). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.17    Transmembrane    17 -  33 (  14 -  39)
   INTEGRAL    Likelihood = -3.77    Transmembrane    92 - 108 (  88 - 108)
   INTEGRAL    Likelihood = -2.87    Transmembrane   144 - 160 ( 142 - 162)
   INTEGRAL    Likelihood = -1.65    Transmembrane    42 -  58 (  42 -  59)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10293> which encodes amino acid sequence <SEQ ID 10294> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3163> which encodes the amino acid sequence <SEQ ID 3164>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.64    Transmembrane    76 -  92 (  72 - 102)
   INTEGRAL    Likelihood = -5.36    Transmembrane   136 - 152 ( 131 - 164)
   INTEGRAL    Likelihood = -2.34    Transmembrane    98 - 114 (  97 - 114)

----- Final Results -----
               bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 145/180 (80%), Positives = 160/180 (88%)

Query: 8    MMKKENIPNLLTVVRILMIPLFIVLTSVTTSTTWHIVAAIVFAIASLTDYLDGYLARKWQ 67
            M+KKENIPNLLT+VRI MIP F+ +TS +    WHI AA++FAIAS TDYLDGYLARKW
Sbjct: 1    MIKKENIPNLLTLVRIAMIPFFLFITSSSNKVGWHIFAAVIFAIASFTDYLDGYLARKWH 60
```

```
Query: 68  VVTNFGKFADPLADKMLVMSAFIMLVGLDLAPAWVSAIIICRELAVTGLRLLLVETGGTV 127
           V +NFGKFADPLADKMLVMSAFIMLVGL L PAWVSA+IICRELAVTGLRLLLVETGG V
Sbjct: 61  VASNFGKFADPLADKMLVMSAFIMLVGLGLVPAWVSAVIICRELAVTGLRLLLVETGGKV 120

Query: 128 LAAAMPGKIKTATQMFAVIFLLVHWMTLGNIMLYIALFFTLYSGYDYFKGAGFLFKDTFK 187
           LAAAMPGKIKTATQM ++I LL HW+ LGN++LYIALFFT+YSGYDYFKGA FLFKDTFK
Sbjct: 121 LAAAMPGKIKTATQMLSIILLLCHWIFLGNVLLYIALFFTIYSGYDYFKGASFLFKDTFK 180
```

A related GBS gene <SEQ ID 8705> and protein <SEQ ID 8706> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 4
SRCFLG: 0
McG: Length  of  UR:    9
     Peak Value of UR:   3.03
     Net Charge of CR: 1
McG: Discrim Score:      6.36
GvH: Signal Score (-7.5): -0.400001
     Possible site: 48
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 49
ALOM program   count: 2 value:  -3.77 threshold:  0.0
   INTEGRAL    Likelihood = -3.77   Transmembrane   85 - 101 (  81 - 101)
   INTEGRAL    Likelihood = -2.87   Transmembrane  137 - 153 ( 135 - 155)
   PERIPHERAL  Likelihood =  1.27     109
 modified ALOM score:   1.25
icml HYPID: 7  CFP: 0.251

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1027

A DNA sequence (GBSx1097) was identified in *S.agalactiae* <SEQ ID 3165> which encodes the amino acid sequence <SEQ ID 3166>. This protein is predicted to be ABC transporter ATP-binding protein (potA). Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1805(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC61484 GB:AF082738 ABC transporter ATP-binding protein
           [Streptococcus pyogenes]
 Identities = 201/279 (72%), Positives = 231/279 (82%)

Query: 1   MTNIITVNNLFFKYDSNQTHYQLENVSFHVKQGEWLSIIGHNGSGKSTTVRLIDGLLEAE 60
           M+ II +  + F Y  +Q    L+ VSFHVKQGEWLSIIGHNGSGKSTT+RLIDGLLE E
Sbjct: 18  MSAIIELKKVTFNYHKDQEKPTLDGVSFHVKQGEWLSIIGHNGSGKSTTIRLIDGLLEPE 77

Query: 61  SGQIIIDGQELTEDNVWELRHKIGMVFQNPDNQFVGATVEDDVAFGLENKGIPLKDMKER 120
           SG II+DG  LT  NVWE+RHKIGMVFQNPDNQFVGATVEDDVAFGLENKGI  +D+KER
```

```
Sbjct: 78  SGSIIVDGDLLTITNVWEIRHKIGMVFQNPDNQFVGATVEDDVAFGLENKGIAHEDIKER 137

Query: 121 VDQALDLVGMSEFKMREPARLSGGQKQRVAIAGAVAMRPQVIILDEATSMLDPEGRLELI 180
           V+ AL+LVGM  FK +EPARLSGGQKQRVAIAGAVAM+P++IILDEATSMLDP+GRLELI
Sbjct: 138 VNHALELVGMQNFKEKEPARLSGGQKQRVAIAGAVAMKPKIIILDEATSMLDPKGRLELI 197

Query: 181 RTIRAIRQKYNLTVISITHDLDEVALSDRVIVMKNGKVESTSTPKALFGRGNRLISLGLD 240
           +TI+ IR  Y LTVISITHDLDEVALSDRV+VMK+G+VESTSTP+ LF RG+ L+ LGLD
Sbjct: 198 KTIKNIRDDYQLTVISITHDLDEVALSDRVLVMKDGQVESTSTPEQLFARGDELLQLGLD 257

Query: 241 VPFTSRLMAELAANGLDIGTEYLTEKELEEQLWELNLKM 279
           +PFT+ ++  L   G  I   YLTEKELE QL +L  KM
Sbjct: 258 IPFTTSVVQMLQEEGYPIDYGYLTEKELENQLCQLISKM 296
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3167> which encodes the amino acid sequence <SEQ ID 3168>. Analysis of this protein sequence reveals the following:

```
     Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2235(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 247-249
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 200/279 (71%), Positives = 231/279 (82%)

Query: 1   MTNIITVNNLFFKYDSNQTHYQLENVSFHVKQGEWLSIIGHNGSGKSTTVRLIDGLLEAE 60
           M+ II +  + F Y  +Q    L+ VSFHVKQGEWLSIIGHNGSGKSTT+RLIDGLLE E
Sbjct: 18  MSAIIELKKVTFNYHKDQEKPTLDGVSFHVKQGEWLSIIGHNGSGKSTTIRLIDGLLEPE 77

Query: 61  SGQIIIDGQELTEDNVWELRHKIGMVFQNPDNQFVGATVEDDVAFGLENKGIPLKDMKER 120
           SG II+DG  LT  NVWE+RHKIGMVFQNPDNQFVGATVEDDVAFGLENKGI  +D+KER
Sbjct: 78  SGSIIVDGDLLTITNVWEIRHKIGMVFQNPDNQFVGATVEDDVAFGLENKGIAHEDIKER 137

Query: 121 VDQALDLVGMSEFKMREPARLSGGQKQRVAIAGAVAMRPQVIILDEATSMLDPEGRLELI 180
           V+ AL+LVGM  FK +EPARLSGGQKQRVAIAGAVAM+P++IILDEATSMLDP+GRLELI
Sbjct: 138 VNHALELVGMQNFKEKEPARLSGGQKQRVAIAGAVAMKPKIIILDEATSMLDPKGRLELI 197

Query: 181 RTIRAIRQKYNLTVISITHDLDEVALSDRVIVMKNGKVESTSTPKALFGRGNRLISLGLD 240
           +TI+ IR  Y LTVISITHDLDEVALSDRV+VMK+G+VESTSTP+ LF RG+ L+ LGLD
Sbjct: 198 KTIKNIRDDYQLTVISITHDLDEVALSDRVLVMKDGQVESTSTPEQLFARGDELLQLGLD 257

Query: 241 VPFTSRLMAELAANGLDIGTEYLTEKELEEQLWELNLKM 279
           +PFT+ ++  L   G  +   YLTEKELE QL +L  KM
Sbjct: 258 IPFTTSVVQMLQEEGYPVDYGYLTEKELENQLCQLISKM 296
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1028

A DNA sequence (GBSx1098) was identified in *S.agalactiae* <SEQ ID 3169> which encodes the amino acid sequence <SEQ ID 3170>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane  154 - 170 ( 154 - 170)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11922 GB:Z99104 similar to ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 141/242 (58%), Positives = 188/242 (77%), Gaps = 1/242 (0%)

Query: 16  TPFEGRALFDVNLKIEDASYTAFIGHTGSGKSTIMQLLNGLHIPTKGEVIVDDFSIKAGD 75
           TPFE  AL+D+N  I++ SY A IGHTGSGKST++Q LNGL  PTKG++ +    I+AG
Sbjct: 3   TPFERLALYDINASIKEGSYVAVIGHTGSGKSTLLQHLNGLLKPTKGQISLGSTVIQAGK 62

Query: 76  KNKEIKFIRQKVGLVFQFPESQLFEETVLKDVAFGPQNFGISQIEAERLAEEKLRLVGIS 135
           KNK++K +R+KVG+VFQFPE QLFEETVLKD++FGP NFG+ + +AE+ A E L+LVG+S
Sbjct: 63  KNKDLKKLRKKVGIVFQFPEHQLFEETVLKDISFGPMNFGVKKEDAEQKAREMLQLVGLS 122

Query: 136 EDLFDKNPFELSGGQMRRVAIAGILAMEPKVLVLDEPTAGLDPKGRKELMTLFKNLHKKG 195
           E+L D++PFELSGGQMRRVAIAG+LAM+P+VLVLDEPTAGLDP+GRKE+M +F  LH++G
Sbjct: 123 EELLDRSPFELSGGQMRRVAIAGVLAMDPEVLVLDEPTAGLDPRGRKEIMDMFYELHQRG 182

Query: 196 -MTIVLVTHLMDDVADYADYVYVLEAGKVTLSGQPKQIFQEVELLESKQLGVPKITKFAQ 254
            +T +LVTH M+D A YAD + V+  G +  SG P+ +F + E +    L +P+  KF +
Sbjct: 183 NLTTILVTHSMEDAAAYADEMIVMHKGTIQASGSPRDLFLKGEEMAGWGLDLPETIKFQR 242

Query: 255 RL 256
            L
Sbjct: 243 HL 244
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3171> which encodes the amino acid sequence <SEQ ID 3172>. Analysis of this protein sequence reveals the following:

```
      Possible site: 40

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane  154 - 170 ( 154 - 170)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB11922 GB:Z99104 similar to ABC transporter (ATP-binding
          protein) [Bacillus subtilis]
 Identities = 146/259 (56%), Positives = 187/259 (71%), Gaps = 2/259 (0%)

Query: 16  TPFEGRALFNINLDILDGSYTAFIGHTGSGKSTIMQLLNGLHVPTTGIVSVDKQDITNHS 75
           TPFE  AL++IN  I +GSY A IGHTGSGKST++Q LNGL  PT G +S+    I
Sbjct: 3   TPFERLALYDINASIKEGSYVAVIGHTGSGKSTLLQHLNGLLKPTKGQISLGSTVIQAGK 62

Query: 76  KNKEIKSIRKHVGLVFQFPESQLFEETVLKDVAFGPQNFGVSPEEAEALAREKLALVGIS 135
           KNK++K +RK VG+VFQFPE QLFEETVLKD++FGP NFGV  E+AE  ARE L LVG+S
Sbjct: 63  KNKDLKKLRKKVGIVFQFPEHQLFEETVLKDISFGPMNFGVKKEDAEQKAREMLQLVGLS 122

Query: 136 ENLFEKNPFELSGGQMRRVAIAGILAMQPKVLVLDEPTAGLDPKGRKELMTIFKKLHQSG 195
           E L +++PFELSGGQMRRVAIAG+LAM P+VLVLDEPTAGLDP+GRKE+M +F +LHQ G
Sbjct: 123 EELLDRSPFELSGGQMRRVAIAGVLAMDPEVLVLDEPTAGLDPRGRKEIMDMFYELHQRG 182

Query: 196 -MTIVLVTHLMDDVANYADFVYVLDKGKIILSGKPKTIFQQVSLLEKKQLGVPKVTKLAQ 254
            +T +LVTH M+D A YAD + V+ KG I  SG P+ +F +    +    L +P+  K  +
Sbjct: 183 NLTTILVTHSMEDAAAYADEMIVMHKGTIQASGSPRDLFLKGEEMAGWGLDLPETIKFQR 242

Query: 255 RL-VDRGIPISSLPITLEE 272
            L    G+  +   +T+E+
Sbjct: 243 HLEAALGVRFNEPMLTIED 261
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 218/280 (77%), Positives = 241/280 (85%)

Query: 1   MGIEFKNVSYTYQAGTPFEGRALFDVNLKIEDASYTAFIGHTGSGKSTIMQLLNGLHIPT 60
            M I  +NVSYTYQAGTPFEGRALF++NL I D SYTAFIGHTGSGKSTIMQLLNGLH+PT
Sbjct: 1   MSINLQNVSYTYQAGTPFEGRALFNINLDILDGSYTAFIGHTGSGKSTIMQLLNGLHVPT 60

Query: 61  KGEVIVDDFSIKAGDKNKEIKFIRQKVGLVFQFPESQLFEETVLKDVAFGPQNFGISQIE 120
             G V VD   I    KNKEIK IR+ VGLVFQFPESQLFEETVLKDVAFGPQNFG+S  E
Sbjct: 61  TGIVSVDKQDITNHSKNKEIKSIRKHVGLVFQFPESQLFEETVLKDVAFGPQNFGVSPEE 120

Query: 121 AERLAEEKLRLVGISEDLFDKNPFELSGGQMRRVAIAGILAMEPKVLVLDEPTAGLDPKG 180
            AE LA EKL LVGISE+LF+KNPFELSGGQMRRVAIAGILAM+PKVLVLDEPTAGLDPKG
Sbjct: 121 AEALAREKLALVGISENLFEKNPFELSGGQMRRVAIAGILAMQPKVLVLDEPTAGLDPKG 180

Query: 181 RKELMTLFKNLHKKGMTIVLVTHLMDDVADYADYVYVLEAGKVTLSGQPKQIFQEVELLE 240
            RKELMT+FK LH+ GMTIVLVTHLMDDVA+YAD+VYVL+ GK+ LSG+PK IFQ+V LLE
Sbjct: 181 RKELMTIFKKLHQSGMTIVLVTHLMDDVANYADFVYVLDKGKIILSGKPKTIFQQVSLLE 240

Query: 241 SKQLGVPKITKFAQRLSHKGLNLPSLPITINEFVEAIKHG 280
             KQLGVPK+TK AQRL  +G+ + SLPIT+ E  E +KHG
Sbjct: 241 KKQLGVPKVTKLAQRLVDRGIPISSLPITLEELREVLKHG 280
```

SEQ ID 3170 (GBS401) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 3; MW 34.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 4; MW 59kDa).

GBS401-GST was purified as shown in Figure 218, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1029

A DNA sequence (GBSx1099) was identified in *S.agalactiae* <SEQ ID 3173> which encodes the amino acid sequence <SEQ ID 3174>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.46   Transmembrane   47 -  63 (  25 -  69)
   INTEGRAL    Likelihood = -8.81   Transmembrane  252 - 268 ( 249 - 269)
   INTEGRAL    Likelihood = -7.91   Transmembrane  116 - 132 ( 110 - 141)
   INTEGRAL    Likelihood = -4.25   Transmembrane   29 -  45 (  25 -  46)
   INTEGRAL    Likelihood = -2.55   Transmembrane   77 -  93 (  77 -  95)
   INTEGRAL    Likelihood = -0.43   Transmembrane  199 - 215 ( 199 - 215)

----- Final Results -----
               bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8707> which encodes amino acid sequence <SEQ ID 8708> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
SRCFLG: 0
McG: Length  of  UR:   8
     Peak Value of UR:   0.65
     Net Charge of CR: 1
McG: Discrim Score:    -10.55
GvH: Signal Score (-7.5): 1.45
```

```
     Possible site: 37
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program  count: 6 value: -10.46 threshold:  0.0
   INTEGRAL    Likelihood =-10.46   Transmembrane   41 -  57 (  19 -  63)
   INTEGRAL    Likelihood = -8.81   Transmembrane  246 - 262 ( 243 - 263)
   INTEGRAL    Likelihood = -7.91   Transmembrane  110 - 126 ( 104 - 135)
   INTEGRAL    Likelihood = -4.25   Transmembrane   23 -  39 (  19 -  40)
   INTEGRAL    Likelihood = -2.55   Transmembrane   71 -  87 (  71 -  89)
   INTEGRAL    Likelihood = -0.43   Transmembrane  193 - 209 ( 193 - 209)
   PERIPHERAL  Likelihood =  0.79      90
 modified ALOM score:   2.59
icm1 HYPID: 7  CFP: 0.518

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11923 GB:Z99104 ybaF [Bacillus subtilis]
 Identities = 133/263 (50%), Positives = 191/263 (72%)

Query: 7   MDKLILGRYIPGNSLIHKLDPRSKLLAMLLFIIIVFWANNVVTNVIVFIFTLVIVGLSQI 66
           MD +I+G+Y+PG SL+H+LDPR+KL+ + LF+ IVF ANNV T  ++ +FT+ +V L+++
Sbjct: 2   MDSMIIGKYVPGTSLVHRLDPRTKLITIFLFVCIVFLANNVQTYALLGLFTIGVVSLTRV 61

Query: 67  KFSYFFNGIKPMVGIILFTTLFQMLFAQGGQVIFSFWIFSITSLGLQQAALIFMRFVLII 126
            FS+   G+KP++ I+LFT L  +L    G +IF      +   GL Q   I +RFV +I
Sbjct: 62  PFSFLMKGLKPIIWIVLFTFLLHILMTHEGPIIFQIGFSRVYEGGLVQGIFISLRFVYLI 121

Query: 127 FFSTLLTLTTTPLSLADAVESLLKPLEVLRVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 186
             +TLLTLTTTP+ + D +E LL PL+ L++P HE+ LM+S+SLRF+PTLM++T +IM A
Sbjct: 122 LITTLLTLTTTPIEITDGMEQLLNPLKKLKLPVHELALMMSISLRFIPTLMEETDKIMKA 181

Query: 187 QRARGVDFGEGNLIHKVKSIIPILIPLFASSFKRADALAIAMEARGYQGGANRSKYRLLK 246
           Q ARGVDF  G +  +VK+I+P+L+PLF S+FKRA+ LA+AMEARGYQGG  R+KYR L
Sbjct: 182 QMARGVDFTSGPVKERVKAIVPLLVPLFVSAFKRAEELAVAMEARGYQGGEGRTKYRKLV 241

Query: 247 WTVRDTFSILLMLLLGLSLFLLK 269
           WT +DT  I+ +++L   LF L+
Sbjct: 242 WTGKDTSVIVSLIVLAALLFSLR 264
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3175> which encodes the amino acid sequence <SEQ ID 3176>. Analysis of this protein sequence reveals the following:

```
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.50   Transmembrane  246 - 262 ( 243 - 265)
   INTEGRAL    Likelihood = -9.34   Transmembrane  110 - 126 ( 103 - 135)
   INTEGRAL    Likelihood = -6.69   Transmembrane   41 -  57 (  40 -  58)
   INTEGRAL    Likelihood = -2.81   Transmembrane   23 -  39 (  21 -  40)
   INTEGRAL    Likelihood = -1.01   Transmembrane   62 -  78 (  62 -  78)
   INTEGRAL    Likelihood = -0.27   Transmembrane  193 - 209 ( 193 - 209)

----- Final Results -----
              bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB11923 GB:Z99104 ybaF [Bacillus subtilis]
 Identities = 138/263 (52%), Positives = 195/263 (73%)
```

```
Query: 1   MDKLILGRYIPGDSLIHRLDPRSKLLAMIIYIVIIFWANNVVTNLLMLTFTLAVVFLSKI 60
           MD +I+G+Y+PG SL+HRLDPR+KL+ + +++ I+F ANNV T  L+  FT+ VV L+++
Sbjct: 2   MDSMIIGKYVPGTSLVHRLDPRTKLITIFLFVCIVFLANNVQTYALLGLFTIGVVSLTRV 61

Query: 61  KLSFFLNGVKPMIGIILFTTLFQMFFSQGGKVIFSWWFISITDLGLSQAILIFMRFVLII 120
             SF + G+KP+I I+LFT L  +  +  G +IF   F  + + GL Q I I +RFV +I
Sbjct: 62  PFSFLMKGLKPIIWIVLFTFLLHILMTHEGPIIFQIGFSRVYEGGLVQGIFISLRFVYLI 121

Query: 121 FFSTLLTLTTTPLSLSDAVESLLKPLTRFKVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 180
             +TLLTLTTTP+ ++D +E LL PL + K+P HE+ LM+S+SLRF+PTLM++T +IM A
Sbjct: 122 LITTLLTLTTTPIEITDGMEQLLNPLKKLKLPVHELALMMSISLRFIPTLMEETDKIMKA 181

Query: 181 QRARGVDFGEGNLIQKVKSIIPILIPLFASSFKRADALAIAMEARGYQGGEGRTKYRQLD 240
           Q ARGVDF  G + ++VK+I+P+L+PLF S+FKRA+ LA+AMEARGYQGGEGRTKYR+L
Sbjct: 182 QMARGVDFTSGPVKERVKAIVPLLVPLFVSAFKRAEELAVAMEARGYQGGEGRTKYRKLV 241

Query: 241 WQLKDSLAIGIVSLLGLLLFFLK 263
           W  KD+  I  + +L  LLF L+
Sbjct: 242 WTGKDTSVIVSLIVLAALLFSLR 264
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 210/263 (79%), Positives = 237/263 (89%)

Query: 7   MDKLILGRYIPGNSLIHKLDPRSKLLAMLLFIIIVFWANNVVTNVIVFIFTLVIVGLSQI 66
           MDKLILGRYIPG+SLIH+LDPRSKLLAM+++I+I+FWANNVVTN+++  FTL +V LS+I
Sbjct: 1   MDKLILGRYIPGDSLIHRLDPRSKLLAMIIYIVIIFWANNVVTNLLMLTFTLAVVFLSKI 60

Query: 67  KFSYFFNGIKPMVGIILFTTLFQMLFAQGGQVIFSFWIFSITSLGLQQAALIFMRFVLII 126
           K S+F NG+KPM+GIILFTTLFQM F+QGG+VIFS+W  SIT LGL QA LIFMRFVLII
Sbjct: 61  KLSFFLNGVKPMIGIILFTTLFQMFFSQGGKVIFSWWFISITDLGLSQAILIFMRFVLII 120

Query: 127 FFSTLLTLTTTPLSLADAVESLLKPLEVLRVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 186
           FFSTLLTLTTTPLSL+DAVESLLKPL   +VPAHEIGLMLSLSLRFVPTLMDDTTRIMNA
Sbjct: 121 FFSTLLTLTTTPLSLSDAVESLLKPLTRFKVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 180

Query: 187 QRARGVDFGEGNLIHKVKSIIPILIPLFASSFKRADALAIAMEARGYQGGANRSKYRLLK 246
           QRARGVDFGEGNLI KVKSIIPILIPLFASSFKRADALAIAMEARGYQGG  R+KYR L
Sbjct: 181 QRARGVDFGEGNLIQKVKSIIPILIPLFASSFKRADALAIAMEARGYQGGEGRTKYRQLD 240

Query: 247 WTVRDTFSILLMLLLGLSLFLLK 269
           W ++D+ +I ++ LLGL LF LK
Sbjct: 241 WQLKDSLAIGIVSLLGLLLFFLK 263
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1030

A DNA sequence (GBSx1101) was identified in *S.agalactiae* <SEQ ID 3179> which encodes the amino acid sequence <SEQ ID 3180>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05   Transmembrane   22 -  38 (  16 -  43)

----- Final Results -----
            bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3181> which encodes the amino acid sequence <SEQ ID 3182>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 116/233 (49%), Positives = 140/233 (59%), Gaps = 39/233 (16%)

Query: 9   KLNVKKHHLAYGAITLVALFSCILAVMVIFKSSQVTTESLSKADKVRVAKKSK------- 61
           K N+K+ +  +G   LVAL   ILA++  F S    T+S +K  + ++    K
Sbjct: 4   KENLKQRYFNFG---LVALALTILAIIFAFSSKNADTKSYAKKSESKMVTIDKAPKNNHA 60

Query: 62  MTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVTASEEAAVEQAVVTENTP 121
           +TK  SK K + +   P P+  ++ AP               T +EE  V Q  VT
Sbjct: 61  ITKEESKEKAKSIASEPIPTVENSVAP---------------TVTEEVPVVQQEVT---- 101

Query: 122 ATSQAQQAYAVTETTYRPAQHQTSGQVLSNGNTAGAIGSAAAAQMAAATGVPQSTWEHII 181
                Q    V+   Y P     +  VLSNGNTAG +GS AAAQMAAATGVPQSTWEHII
Sbjct: 102 -----QTVQQVSSVAYNP-----NNVVLSNGNTAGIVGSQAAAQMAAATGVPQSTWEHII 151

Query: 182 ARESNGNPNVANASGASGLFQTMPGWGSTATVQDQVNSAIKAYRAQGLSAWGY 234
           ARESNGNPN ANASGASGLFQTMPGWGSTATV+DQVN+A+KAY AQGLSAWGY
Sbjct: 152 ARESNGNPNAANASGASGLFQTMPGWGSTATVEDQVNAALKAYSAQGLSAWGY 204
```

A related GBS gene <SEQ ID 8713> and protein <SEQ ID 8714> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      2.48
GvH: Signal Score (-7.5): -3.74
     Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -12.05 threshold:  0.0
   INTEGRAL    Likelihood =-12.05   Transmembrane   22 -  38 (  16 -  43)
   PERIPHERAL  Likelihood =  4.29     156
 modified ALOM score:   2.91

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
61.8/68.7% over 114aa
                                                                     Staphylococcus aureus
  GP|7959131| secretory protein SAI-B Insert characterized

ORF01057(664 - 1002 of 1302)
GP|7959131|dbj|BAA95959.1||AB042839(119   -   233   of   233)   secretory   protein   SAI-B
{Staphylococcus aureus}
%Match = 15.1
%Identity = 61.7  %Similarity = 68.7
Matches = 71  Mismatches = 34  Conservative Sub.s = 8

438       468       498       528       558       588       618       648
```

```
IFKSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVTASEEAAVEQAVV

VDQAHLVDLAHNHQDQLNAAPIKDGAYDIHFVKDGFQYNFTSNGTTWSWSYEAANGQTAGFSNVAGADYTTSYNQGSNVQ
               50        60        70        80        90       100       110

678       708       735       762       792       822       852       882
TENTPATSQAQQAYAVTETTYRP-AQHQTSGQV-LSNGNTAGAIGSAAAAQMAAATGVPQSTWEHIIARESNGNPNVANA
: :  | |      ||:  ||   :   ||  | ||||||||| ||:||  ||  |||| |||  ||||||||  |  |
SVSYNAQSSNSNVEAVSAPTYHNYSTSTTSSSVRLSNGNTAGATGSSAAQIMAQRTGVPASTWAAIIARESNGQVNAYNP
              130       140       150       160       170       180       190

912       942       972      1002      1032      1062      1092      1122
SGASGLFQTMPGWGSTATVQDQVNSAIKAYRAQGLSAWGY**IAIN*LYTVVNNNYRLLKQINKNATVKL*RFYLFSGKE
|||||||||||||| | ||  |:|:|:|||:|||| |||:
SGASGLFQTMPGWGPTNTVDQQINAAVKAYKAQGLGAWGF
```

SEQ ID 3180 (GBS25) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 5; MW 25kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 11; MW 50kDa), Figure 63 (lane 6; MW 50.3kDa), Figure 66 (lane 6; MW 50kDa) and in Figure 175 (lane 8 & 9; MW 50kDa).

Purified GBS25-GST is shown in Figure 9A, Figure 193 (lane 11) and Figure 210 (lane 5).

The purified GBS25-GST fusion product was used to immunise mice (lane 1+2+3 products; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 95B), FACS (Figure 95C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1031

A DNA sequence (GBSx1103) was identified in *S.agalactiae* <SEQ ID 3183> which encodes the amino acid sequence <SEQ ID 3184>. This protein is predicted to be L-serine dehydratase 1 (sdaA-2). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.85   Transmembrane  205 - 221 ( 205 - 221)
   INTEGRAL    Likelihood = -0.59   Transmembrane  171 - 187 ( 171 - 187)
   INTEGRAL    Likelihood = -0.53   Transmembrane  226 - 242 ( 226 - 242)

----- Final Results -----
               bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13459 GB:Z99112 similar to L-serine dehydratase [Bacillus subtilis]
 Identities = 176/289 (60%), Positives = 224/289 (76%), Gaps = 1/289 (0%)

Query: 1    MFYTIEELVEQANSQHKGNIAELMIQTEIEMTGRSREEIRYIMSRNLEVMKASVIDGLTP 60
            MF  ++EL+E    + +  I+++MI  E+E+T +++E+I   M  NL VM+A+V  GL
Sbjct: 1    MFRNVKELIE-ITKEKQILISDVMIAQEMEVTEKTKEDIFQQMDHNLSVMEAAVQKGLEG 59

Query: 61   SKSISGLTGGDAVKMDQYLQSGKTISDTTILAAVRNAMAVNELNAKMGLVCATPTAGSAG 120
              S +GLTGGDAVK+  Y++SGK++S   IL AV  A+A NE+NA MG +CATPTAGSAG
Sbjct: 60   VTSQTGLTGGDAVKLQAYIRSGKSLSGPLILDAVSKAVATNEVNAAMGTICATPTAGSAG 119
```

```
Query: 121 CLPAVISTAIEKLNLTEEEQLDFLFTAGAFGLVIGNNASISGAEGGCQAEVGSASAMAAA 180
            +P  +    EKLN T E+ + FLFTAGAFG V+ NNASISGA GGCQAEVGSAS MAAA
Sbjct: 120 VVPGTLFAVKEKLNPTREQMIRFLFTAGAFGFVVANNASISGAAGGCQAEVGSASGMAAA 179

Query: 181 ALVMAAGGTPFQASQAIAFVIKNMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMALA 240
           A+V  AGGTP Q+++A+A  +KNMLGL+CDPVAGLVEVPCVKRNA+G+S A++AADMALA
Sbjct: 180 AIVEMAGGTPEQSAEAMAITLKNMLGLVCDPVAGLVEVPCVKRNAMGASNAMIAADMALA 239

Query: 241 GIESQIPVDEVIDAMYQVGSSLPTAFRETAEGGLAATPTGRRYSKEIFG 289
           GI S+IP DEVIDAMY++G ++PTA RET +GGLAATPTGR   K+IFG
Sbjct: 240 GITSRIPCDEVIDAMYKIGQTMPTALRETGQGGLAATPTGRELEKKIFG 288
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3185> which encodes the amino acid sequence <SEQ ID 3186>. Analysis of this protein sequence reveals the following:

```
     Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  196 - 212 ( 196 - 213)
   INTEGRAL    Likelihood = -0.27   Transmembrane  226 - 242 ( 226 - 242)

----- Final Results -----
               bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13459 GB:Z99112 similar to L-serine dehydratase [Bacillus subtilis]
 Identities = 173/289 (59%), Positives = 222/289 (75%), Gaps = 1/289 (0%)

Query: 1   MFYTIEELVKQADQQFNGNIAELMIATEVEMSGRNREDIIKIMSRNLQVMKAAVTEGLTS 60
           MF  ++EL++   ++     I+++MIA E+E++ + +EDI + M  NL VM+AAV +GL
Sbjct: 1   MFRNVKELIEITKEK-QILISDVMIAQEMEVTEKTKEDIFQQMDHNLSVMEAAVQKGLEG 59

Query: 61  TKSISGLTGGDAVKMDNYIKKGNSLSDTTILNAVRNAIAVNELNAKMGLVCATPTAGSAG 120
             S +GLTGGDAVK+  YI+ G SLS   IL+AV  A+A NE+NA MG +CATPTAGSAG
Sbjct: 60  VTSQTGLTGGDAVKLQAYIRSGKSLSGPLILDAVSKAVATNEVNAAMGTICATPTAGSAG 119

Query: 121 CLPAVLATAIEKLDLSEKEQLEFLFTAGAFGLVIGNNASISGAEGGCQAEVGSAAAMSAA 180
            +P  L    EKL+ + ++ + FLFTAGAFG V+ NNASISGA GGCQAEVGSA+ M+AA
Sbjct: 120 VVPGTLFAVKEKLNPTREQMIRFLFTAGAFGFVVANNASISGAAGGCQAEVGSASGMAAA 179

Query: 181 ALVKAAGGTSHQASQAIAFVIKNLLGLVCDPVAGLVEVPCVKRNALGASFALVAADMALA 240
           A+V+ AGGT  Q+++A+A  +KN+LGLVCDPVAGLVEVPCVKRNA+GAS A++AADMALA
Sbjct: 180 AIVEMAGGTPEQSAEAMAITLKNMLGLVCDPVAGLVEVPCVKRNAMGASNAMIAADMALA 239

Query: 241 DIDSQIPVDEVIDAMYQVGSAMPTAFRETAEGGLAATPTGRRYSVEIFG 289
            I S+IP DEVIDAMY++G  MPTA RET +GGLAATPTGR    +IFG
Sbjct: 240 GITSRIPCDEVIDAMYKIGQTMPTALRETGQGGLAATPTGRELEKKIFG 288
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 244/290 (84%), Positives = 273/290 (94%)

Query: 1   MFYTIEELVEQANSQHKGNIAELMIQTEIEMTGRSREEIRYIMSRNLEVMKASVIDGLTP 60
           MFYTIEELV+QA+ Q  GNIAELMI TE+EM+GR+RE+I  IMSRNL+VMKA+V +GLT
Sbjct: 1   MFYTIEELVKQADQQFNGNIAELMIATEVEMSGRNREDIIKIMSRNLQVMKAAVTEGLTS 60

Query: 61  SKSISGLTGGDAVKMDQYLQSGKTISDTTILAAVRNAMAVNELNAKMGLVCATPTAGSAG 120
           +KSISGLTGGDAVKMD Y++ G ++SDTTIL AVRNA+AVNELNAKMGLVCATPTAGSAG
Sbjct: 61  TKSISGLTGGDAVKMDNYIKKGNSLSDTTILNAVRNAIAVNELNAKMGLVCATPTAGSAG 120

Query: 121 CLPAVISTAIEKLNLTEEEQLDFLFTAGAFGLVIGNNASISGAEGGCQAEVGSASAMAAA 180
           CLPAV++TAIEKL+L+E+EQL+FLFTAGAFGLVIGNNASISGAEGGCQAEVGSA+AM+AA
Sbjct: 121 CLPAVLATAIEKLDLSEKEQLEFLFTAGAFGLVIGNNASISGAEGGCQAEVGSAAAMSAA 180

Query: 181 ALVMAAGGTPFQASQAIAFVIKNMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMALA 240
```

```
           ALV AAGGT  QASQAIAFVIKN+LGL+CDPVAGLVEVPCVKRNALG+SFALVAADMALA
Sbjct: 181 ALVKAAGGTSHQASQAIAFVIKNLLGLVCDPVAGLVEVPCVKRNALGASFALVAADMALA 240

Query: 241 GIESQIPVDEVIDAMYQVGSSLPTAFRETAEGGLAATPTGRRYSKEIFGE 290
            I+SQIPVDEVIDAMYQVGS++PTAFRETAEGGLAATPTGRRYS EIFGE
Sbjct: 241 DIDSQIPVDEVIDAMYQVGSAMPTAFRETAEGGLAATPTGRRYSVEIFGE 290
```

SEQ ID 3184 (GBS358) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 6; MW 35kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1032

A DNA sequence (GBSx1104) was identified in *S.agalactiae* <SEQ ID 3187> which encodes the amino acid sequence <SEQ ID 3188>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06216 GB:AP001515 L-serine dehydratase beta subunit [Bacillus halodurans]
 Identities = 101/216 (46%), Positives = 156/216 (71%), Gaps = 2/216 (0%)

Query: 4   LKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVHSIFGE-PSEVTFHLYNSFAKTYQGHGT 62
           +K+++VFDIIGPVMIGPSSSHTAGA RIG+V  ++FG+ P     + Y SFA+TY+GHGT
Sbjct: 1   MKYRTVFDIIGPVMIGPSSSHTAGAARIGRVARTLFGQQPERCDIYFYGSFAETYKGHGT 60

Query: 63  DKALVAGILGMDTDNPDIKNSLEIAHQKGIKIYWDILKDSNSPHPNTAKITVKNGDRSMS 122
           D A+V GIL  DT +P I  SL++A +KG+++Y+   +++ + HPNTAK+ ++ G+  +
Sbjct: 61  DVAIVGGILDFDTFDPRIPRSLQLAKEKGVRVYFHE-EEAITDHPNTAKVVLQKGEDQLE 119

Query: 123 ITGVSIGGGNIQVTELNGFSVSLTMNTPTLIIVHQDIPGMIAKVTDILSDFNINIAQMNV 182
           + GVSIGGG I++ ELNGF + L+ N P +++VH D  G+IA V+++L+   INI  M V
Sbjct: 120 VVGVSIGGGKIEIVELNGFHLKLSGNHPAILVVHTDRFGVIASVSNMLAKHEINIGHMEV 179

Query: 183 TRESAGEKAIMIIEVDSRDCQQAVKKIEAIPHLHNV 218
           +R+  G++A+M+IEVD       ++++E +P++  V
Sbjct: 180 SRKEKGKEALMVIEVDQNVDDLLLQELERLPNIVTV 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3189> which encodes the amino acid sequence <SEQ ID 3190>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9161> which encodes the amino acid sequence <SEQ ID 9162>. Analysis of this protein sequence reveals the following:

```
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
```

```
----- Final Results -----
             bacterial outside --- Certainty= 0.300(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 187/223 (83%), Positives = 205/223 (91%), Gaps = 1/223 (0%)

Query: 1   MKHLKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVHSIFGE-PSEVTFHLYNSFAKTYQG 59
           M   KFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVHSIFG+ P EVTFHLYNSFAKTY+G
Sbjct: 3   MNTQKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVHSIFGDIPDEVTFHLYNSFAKTYRG 62

Query: 60  HGTDKALVAGILGMDTDNPDIKNSLEIAHQKGIKIYWDILKDSNSPHPNTAKITVKNGDR 119
           HGTDKALVAGI+GM TDNPDIKNSLEIAHQKGIKIYWDILKDSN+PHPNT KI+VK  D+
Sbjct: 63  HGTDKALVAGIMGMGTDNPDIKNSLEIAHQKGIKIYWDILKDSNAPHPNTVKISVKKADK 122

Query: 120 SMSITGVSIGGGNIQVTELNGFSVSLTMNTPTLIIVHQDIPGMIAKVTDILSDFNINIAQ 179
           ++S+TGVSIGGGNIQVTELNGFSVSL+MNTPT++ VH+DIPGMIAKVTDILS  NINIA
Sbjct: 123 TLSVTGVSIGGGNIQVTELNGFSVSLSMNTPTIVTVHKDIPGMIAKVTDILSSNNINIAT 182

Query: 180 MNVTRESAGEKAIMIIEVDSRDCQQAVKKIEAIPHLHNVNFFD 222
           MNVTRESAGEKA MIIEVDSR+CQ+A  +I  IPH++NVNFFD
Sbjct: 183 MNVTRESAGEKATMIIEVDSRECQEAANQIAKIPHIYNVNFFD 225
```

SEQ ID 3188 (GBS151) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 11; MW 25kDa) and in Figure 165 (lane 14-16; MW 25.3kDa).

The GBS151-GST fusion product was purified (Figure 198, lane 3; Figure 236, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 289), which confirmed that the protein is immunoaccessible on GBS bacteria.

GBS151L was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 8-10; MW 50kDa). GBS151L was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 11 & 12; MW 25kDa), in Figure 128 (lane 7; MW 25kDa) and in Figure 180 (lane 7; MW 25kDa). Purified GBS151L-His is shown in Figure 232 (lanes 5 & 6) and in Figure 240 (lanes 3 & 4).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1033

A DNA sequence (GBSx1105) was identified in *S.agalactiae* <SEQ ID 3191> which encodes the amino acid sequence <SEQ ID 3192>. This protein is predicted to be tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (trmU). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2208(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10291> which encodes amino acid sequence <SEQ ID 10292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04980 GB:AP001511
           (5-methylaminomethyl-2-thiouridylate)-methyltran sferase
           [Bacillus halodurans]
 Identities = 250/359 (69%), Positives = 292/359 (80%), Gaps = 6/359 (1%)

Query: 32  RVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDYKDVAAVADQIG 91
           RVVVGMSGGVDSSVTALLLKEQGYDVIG+FMKNWDDTDE GVCTATEDY+DV  V +Q+G
Sbjct: 10  RVVVGMSGGVDSSVTALLLKEQGYDVIGIFMKNWDDTDENGVCTATEDYQDVVQVCNQLG 69

Query: 92  IPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYAMTLGADYVATG 151
           I YY+VNFEKEYWD+VF YFL EY+AGRTPNPDVMCNKEIKFKAFL++A+TLGADYVATG
Sbjct: 70  IAYYAVNFEKEYWDKVFTYFLEEYKAGRTPNPDVMCNKEIKFKAFLNHALTLGADYVATG 129

Query: 152 HYAQVTRDENGIVHMLRGADNNKDQTYFLSQLSQEQLQKTLFPLGHLQKPEVRRIAEEAG 211
           HYAQV ++ +G    ++RG D NKDQTYFL+ LSQ+QL + +FPLGHL+K EVR IAE AG
Sbjct: 130 HYAQV-KNVDGQYQLIRGKDPNKDQTYFLNALSQQQLSRVMFPLGHLEKKEVRAIAERAG 188

Query: 212 LATAKKKDSTGICFIGEKNFKDFLGQYLPAQPGRMMTVDGRDMGEHAGLMYYTIGQRGGL 271
           LATAKKKDSTGICFIG+++FK+FL  YLPAQPG M T+DG   G H GLMYYT+GQR GL
Sbjct: 189 LATAKKKDSTGICFIGKRDFKEFLSSYLPAQPGEMQTLDGEVKGTHDGLMYYTLGQRQGL 248

Query: 272 GIGGQHGGDNKPWFVVGKDLSKNILYVGQGFYHDSLMSTSLTASEIHFTRDMPNEFKLEC 331
           GI    GG  +PWFV+GK+L KNILYVGQGF+H  L S  L A ++++     ++   EC
Sbjct: 249 GI----GGSGEPWFVIGKNLEKNILYVGQGFHHPGLYSEGLRAIKVNWILRRESDEPFEC 304

Query: 332 TAKFRYRQPDSKVTVYVKGNQA-RVVFDDLQRAITPGQAVVFYNEQECLGGGMIDQAYR 389
           TAKFRYRQPD KVTVY + + A  V+F + QRAITPGQAVVFY+   CLGGG ID   +
Sbjct: 305 TAKFRYRQPDQKVTVYPQSDGAVEVLFAEPQRAITPGQAVVFYDGDVCLGGGTIDHVLK 363
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3193> which encodes the amino acid sequence <SEQ ID 3194>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1691(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 331-333
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04980 GB:AP001511
           (5-methylaminomethyl-2-thiouridylate)-methyltran sferase
           [Bacillus halodurans]
 Identities = 255/359 (71%), Positives = 293/359 (81%), Gaps = 6/359 (1%)

Query: 14  RVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDYKDVAAVADKIG 73
           RVVVGMSGGVDSSVTALLLKEQGYDVIG+FMKNWDDTDE GVCTATEDY+DV  V +++G
Sbjct: 10  RVVVGMSGGVDSSVTALLLKEQGYDVIGIFMKNWDDTDENGVCTATEDYQDVVQVCNQLG 69

Query: 74  IPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYAMTLGADYVATG 133
           I YY+VNFEKEYWD+VF YFL EY+AGRTPNPDVMCNKEIKFKAFL++A+TLGADYVATG
Sbjct: 70  IAYYAVNFEKEYWDKVFTYFLEEYKAGRTPNPDVMCNKEIKFKAFLNHALTLGADYVATG 129

Query: 134 HYAQVKRDENGTVHMLRGADNGKDQTYFLSQLSQEQLQKTLFPLGHLQKSEVREIAERAG 193
           HYAQVK + +G    ++RG D  KDQTYFL+ LSQ+QL + +FPLGHL+K EVR IAERAG
Sbjct: 130 HYAQVK-NVDGQYQLIRGKDPNKDQTYFLNALSQQQLSRVMFPLGHLEKKEVRAIAERAG 188
```

```
Query: 194 LATAKKKDSTGICFIGEKNFKQFLSQYLPAQKGRMMTIDGRDMGEHAGLMYYTIGQRGGL 253
           LATAKKKDSTGICFIG+++FK+FLS YLPAQ G M T+DG   G H GLMYYT+GQR GL
Sbjct: 189 LATAKKKDSTGICFIGKRDFKEFLSSYLPAQPGEMQTLDGEVKGTHDGLMYYTLGQRQGL 248

Query: 254 GIGGQHGGDNQPWFVVGKDLSQNILYVGQGFYHEALMSNSLDASVIHFTREMPEEFTFEC 313
           GI    GG  +PWFV+GK+L +NILYVGQGF+H  L S  L A +++        +  FEC
Sbjct: 249 GI----GGSGEPWFVIGKNLEKNILYVGQGFHHPGLYSEGLRAIKVNWILRRESDEPFEC 304

Query: 314 TAKFRYRQPDSHVAVHVRGDKA-EVVFAEPQRAITPGQAVVFYDGKECLGGGMIDMAYK 371
           TAKFRYRQPD  V V+ + D A EV+FAEPQRAITPGQAVVFYDG  CLGGG ID   K
Sbjct: 305 TAKFRYRQPDQKVTVYPQSDGAVEVLFAEPQRAITPGQAVVFYDGDVCLGGGTIDHVLK 363
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 332/377 (88%), Positives = 349/377 (92%)

Query: 21   GRILMTDNSNIRVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDY 80
            G   MTDNS IRVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDY
Sbjct: 3    GEFFMTDNSKIRVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDY 62

Query: 81   KDVAAVADQIGIPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYA 140
            KDVAAVAD+IGIPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYA
Sbjct: 63   KDVAAVADKIGIPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYA 122

Query: 141  MTLGADYVATGHYAQVTRDENGIVHMLRGADNNKDQTYFLSQLSQEQLQKTLFPLGHLQK 200
            MTLGADYVATGHYAQV RDENG VHMLRGADN KDQTYFLSQLSQEQLQKTLFPLGHLQK
Sbjct: 123  MTLGADYVATGHYAQVKRDENGTVHMLRGADNGKDQTYFLSQLSQEQLQKTLFPLGHLQK 182

Query: 201  PEVRRIAEEAGLATAKKKDSTGICFIGEKNFKDFLGQYLPAQPGRMMTVDGRDMGEHAGL 260
             EVR IAE AGLATAKKKDSTGICFIGEKNFK FL QYLPAQ GRMMT+DGRDMGEHAGL
Sbjct: 183  SEVREIAERAGLATAKKKDSTGICFIGEKNFKQFLSQYLPAQKGRMMTIDGRDMGEHAGL 242

Query: 261  MYYTIGQRGGLGIGGQHGGDNKPWFVVGKDLSKNILYVGQGFYHDSLMSTSLTASEIHFT 320
            MYYTIGQRGGLGIGGQHGGDN+PWFVVGKDLS+NILYVGQGFYH++LMS SL AS IHFT
Sbjct: 243  MYYTIGQRGGLGIGGQHGGDNQPWFVVGKDLSQNILYVGQGFYHEALMSNSLDASVIHFT 302

Query: 321  RDMPNEFKLECTAKFRYRQPDSKVTVYVKGNQARVVFDDLQRAITPGQAVVFYNEQECLG 380
            R+MP EF  ECTAKFRYRQPDS V V+V+G++A VVF + QRAITPGQAVVFY+ +ECLG
Sbjct: 303  REMPEEFTFECTAKFRYRQPDSHVAVHVRGDKAEVVFAEPQRAITPGQAVVFYDGKECLG 362

Query: 381  GGMIDQAYRDDKICQYI 397
            GGMID AY++ + CQYI
Sbjct: 363  GGMIDMAYKNGQPCQYI 379
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1034

A DNA sequence (GBSx1106) was identified in *S.agalactiae* <SEQ ID 3195> which encodes the amino acid sequence <SEQ ID 3196>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.84    Transmembrane  141 - 157 ( 134 - 165)
   INTEGRAL    Likelihood =-11.78    Transmembrane   40 -  56 (  36 -  73)
   INTEGRAL    Likelihood = -4.35    Transmembrane   68 -  84 (  65 -  86)
   INTEGRAL    Likelihood = -3.50    Transmembrane  180 - 196 ( 175 - 199)

----- Final Results -----
               bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15390 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 71/202 (35%), Positives = 120/202 (59%), Gaps = 5/202 (2%)

Query: 1   MISKFILAFMAFFAIMNPISNLPAFMALVADDDQKISRRIAAKGVLLAFVIIVIFVLSGH 60
           M S  +  F++ FA+ NPI N+P F+ L          + IA K  +L+F I+  F++ GH
Sbjct: 2   MFSFIVHVFISLFAVSNPIGNVPIFLTLTEGYTAAERKAIARKAAILSFFILAAFLVFGH 61

Query: 61  LLFNLFGITLAALKISGGILVGIIGYKMINGIHSPTNK-NLEEHKD--DPMNVAVSPLAM 117
           L+F LF I + AL+++GGI +  I Y ++N  S      + +EHK+  +  +++V+PL++
Sbjct: 62  LIFKLFDINIHALRVAGGIFIFGIAYNLLNAKESHVQSLHHDEHKESKEKADISVTPLSI 121

Query: 118 PLLAGPGTIATAMGLSSG--GLSGKLITILAFAILCVIMYVILISANEITKFLGKNAMTI 175
           P++AGPGTIAT M LS+G  G+       ++  A +  + ++     + I+  LGK  M +
Sbjct: 122 PIIAGPGTIATVMSLSAGHSGIGHYAAVMIGIAAVIALTFLFFHYSAFISSKLGKTEMNV 181

Query: 176 ITKMMGLILMTIGIEMLITGIK 197
           IT++MGLIL  + + M+  G+K
Sbjct: 182 ITRLMGLILAVVAVGMIGAGLK 203
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8715> and protein <SEQ ID 8716> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      9.79
GvH: Signal Score (-7.5): -1.53
     Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 4 value: -12.84 threshold:  0.0
   INTEGRAL    Likelihood =-12.84   Transmembrane  141 - 157 ( 134 - 165)
   INTEGRAL    Likelihood =-11.78   Transmembrane   40 -  56 (  36 -  73)
   INTEGRAL    Likelihood = -4.35   Transmembrane   68 -  84 (  65 -  86)
   INTEGRAL    Likelihood = -3.50   Transmembrane  180 - 196 ( 175 - 199)
   PERIPHERAL  Likelihood =  1.27     110
 modified ALOM score:    3.07

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00620(301 - 891 of 1209)
OMNI|NT01BS3953(11 - 212 of 220) conserved hypothetical protein
%Match = 15.8
%Identity = 35.5  %Similarity = 61.5
Matches = 71  Mismatches = 74  Conservative Sub.s = 52

96        126       156       186       216       246       276       306
VQLSSDIVNLTVKLQFT*KVIKQGLCLMIYNEQSHQVKLLFFIMNKNV*AVG*LIRLIVMIKSVNTFN*HLIIK*GNRMI
                                                                              |
                                                                    VQRLSTRRYMMF
                                                                            10

336       366       396       426       456       486       516       546
SKFILAFMAFFAIMNPISNLPAFMALVADDDQKISRRIAAKGVLLAFVIIVIFVLSGHLLFNLFGITLAALKISGGILVG
|  :  |:::||: ||| |:| |: |         : || |  :|:| |:  |:: |||:| || | : ||:::|||::
SFIVHVFISLFAVSNPIGNVPIFLTLTEGYTAAERKAIARKAAILSFFILAAFLVFGHLIFKLFDINIHALRVAGGIFIF
                30        40        50        60        70        80        90

576       603       627       657       687       711       741       771
IIGYKMINGIHSPTNK-NLEEHKD--DPMNVAVSPLAMPLLAGPGTIATAMGLSSG--GLSGKLITILAFAILCVIMYVI
 | | ::|   |     : :|||:  :  :::|:||::|::|||||||| | ||:|  |:      ::  | :  : ::
```

```
GIAYNLLNAKESHVQSLHHDEHKESKEKADISVTPLSIPIIAGPGTIATVMSLSAGHSGIGHYAAVMIGIAAVIALTFLF
               110       120       130       140       150       160       170

801       831       861       891       921       951       981      1011
LISANEITKFLGKNAMTIITKMMGLILMTIGIEMLITGIKIGFHXT*PIPSG*LLKDKC*NKFNXNYDGQSSWNL*VFLT
:  :  |:  |||  | :||::|||||  : : |:  |:|
FHYSAFISSKLGKTEMNVITRLMGLILAVVAVGMIGAGLKGMFPVLTS
               190       200       210       220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1035

A DNA sequence (GBSx1107) was identified in *S.agalactiae* <SEQ ID 3197> which encodes the amino acid sequence <SEQ ID 3198>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1747(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10289> which encodes amino acid sequence <SEQ ID 10290> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45494 GB:U80409 glucose inhibited division protein homolog
           GidA [Lactococcus lactis subsp. cremoris]
 Identities = 394/524 (75%), Positives = 458/524 (87%), Gaps = 2/524 (0%)

Query: 13  KTLLATINLEMLAFMPCNPSIGGSAKGIVVREIDALGGEMGKNIDKTYIQMKMLNTGKGP 72
            KTLL TINL M+AFMPCNPSIGGSAKGIVVREIDALGGEMG+NIDKTYIQMKMLNTGKGP
Sbjct: 12  KTLLMTINLNMVAFMPCNPSIGGSAKGIVVREIDALGGEMGRNIDKTYIQMKMLNTGKGP 71

Query: 73  AVRALRAQADKALYAQTMKQTVEKQENLTLRQAMIDEILVEDGK--VVGVRTATNQKFSA 130
            AVRALRAQADK  YA +MK TV  QENLTLRQ M++E++++D K  V+GVRT+T  ++ A
Sbjct: 72  AVRALRAQADKDEYAASMKNTVSDQENLTLRQGMVEELILDDEKQKVIGVRTSTGTQYGA 131

Query: 131 KSVVITTGTALRGEIILGDLKYSSGPNNSLASVTLADNLRDLGLEIGRFKTGTPPRVKAS 190
            K+V+ITTGTALRGEII+G+LKYSSGPNNSL+S+ LADNLR++G EIGRFKTGTPPRV AS
Sbjct: 132 KAVIITTGTALRGEIIIGELKYSSGPNNSLSSIGLADNLREIGFEIGRFKTGTPPRVLAS 191

Query: 191 SINYEKTEIQPGDEQPNHFSFMSRDEDYITDQVPCWLTYTNTLSHDIINQNLHRAPMFSG 250
            SI+Y+KTEIQPGDE PNHFSFMS DEDY+ DQ+PCWLTYT   SH I+  NLHRAP+FSG
Sbjct: 192 SIDYDKTEIQPGDEAPNHFSFMSSDEDYLKDQIPCWLTYTTENSHTILRDNLHRAPLFSG 251

Query: 251 IVKGVGPRYCPSIEDKIVRFADKERHQLFLEPEGRYTEEVYVQGLSTSLPEDVQVDLLRS 310
            IVKGVGPRYCPSIEDKI RFADK RHQLFLEPEGR TEEVY+ GLSTS+PEDVQ DL++S
Sbjct: 252 IVKGVGPRYCPSIEDKITRFADKPRHQLFLEPEGRNTEEVYIGGLSTSMPEDVQFDLVKS 311

Query: 311 IKGLENAEMMRTGYAIEYDIVLPHQLRATLETKVIAGLFTAGQTNGTSGYEEAAGQGLVA 370
            I GLENA+MMR GYAIEYD+V+PHQLR TLETK+I+GLFTAGQTNGTSGYEEAAGQGLVA
Sbjct: 312 IPGLENAKMMRPGYAIEYDVVMPHQLRPTLETKLISGLFTAGQTNGTSGYEEAAGQGLVA 371

Query: 371 GINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSRAEYRLILRHDNADMR 430
            GINAALK+QGKPE ILKRS+AYIGVMIDDLVTKGTLEPYRLLTSRAEYRLILRHDNAD R
Sbjct: 372 GINAALKIQGKPEFILKRSEAYIGVMIDDLVTKGTLEPYRLLTSRAEYRLILRHDNADRR 431

Query: 431 LTEIGYEIGLVDEERYAIFKKRQMQFENELERLDSIKLKPVSETNKRIQELGFKPLTDAL 490
            LTEIG ++GLV + ++  ++ +  QF+ E++RL+S KLKP+ +T +++ +LGF P+ DAL
Sbjct: 432 LTEIGRQVGLVSDAQWEHYQAKMAQFDREMKRLNSEKLKPLPDTQEKLGKLGFGPIKDAL 491
```

```
Query: 491 TAKEFMRRPQITYAVATDFVGCADEPLDSKVIELLETEIKYEGY 534
           T  EF++RP++ Y    DF+G A E +D  V EL+ETEI YEGY
Sbjct: 492 TGAEFLKRPEVNYDEVIDFIGQAPEVIDRTVSELIETEITYEGY 535
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3199> which encodes the amino acid sequence <SEQ ID 3200>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1064(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 530/610 (86%), Positives = 574/610 (93%)

Query: 1   MEASLAASRMGCKTLLATINLEMLAFMPCNPSIGGSAKGIVVREIDALGGEMGKNIDKTY 60
           +EASLA SRMGCKTLLATINL+MLAFMPCNPSIGGSAKGIVVREIDALGGEMGKNIDKTY
Sbjct: 21  VEASLATSRMGCKTLLATINLDMLAFMPCNPSIGGSAKGIVVREIDALGGEMGKNIDKTY 80

Query: 61  IQMKMLNTGKGPAVRALRAQADKALYAQTMKQTVEKQENLTLRQAMIDEILVEDGKVVGV 120
           IQMKMLNTGKGPAVRALRAQADK+LYA+ MK TVEKQ NLTLRQ MID+ILVEDG+VVGV
Sbjct: 81  IQMKMLNTGKGPAVRALRAQADKSLYAREMKHTVEKQANLTLRQTMIDDILVEDGRVVGV 140

Query: 121 RTATNQKFSAKSVVITTGTALRGEIILGDLKYSSGPNNSLASVTLADNLRDLGLEIGRFK 180
            TAT QKF+AK+VV+TTGTALRGEIILG+LKYSSGPNNSLASVTLADNL+ LGLEIGRFK
Sbjct: 141 LTATGQKFAAKAVVVTTGTALRGEIILGELKYSSGPNNSLASVTLADNLKKLGLEIGRFK 200

Query: 181 TGTPPRVKASSINYEKTEIQPGDEQPNHFSFMSRDEDYITDQVPCWLTYTNTLSHDIINQ 240
           TGTPPRVKASSINY++TEIQPGD++PNHFSFMS+D DY+ DQ+PCWLTYTN  SHDIINQ
Sbjct: 201 TGTPPRVKASSINYDQTEIQPGDDKPNHFSFMSKDADYLKDQIPCWLTYTNQTSHDIINQ 260

Query: 241 NLHRAPMFSGIVKGVGPRYCPSIEDKIVRFADKERHQLFLEPEGRYTEEVYVQGLSTSLP 300
           NL+RAPMFSGIVKGVGPRYCPSIEDKIVRFADKERHQLFLEPEGR TEEVYVQGLSTSLP
Sbjct: 261 NLYRAPMFSGIVKGVGPRYCPSIEDKIVRFADKERHQLFLEPEGRDTEEVYVQGLSTSLP 320

Query: 301 EDVQVDLLRSIKGLENAEMMRTGYAIEYDIVLPHQLRATLETKVIAGLFTAGQTNGTSGY 360
           EDVQ DL+ SIKGLE AEMMRTGYAIEYDIVLPHQLRATLETK+I+GLFTAGQTNGTSGY
Sbjct: 321 EDVQKDLIHSIKGLEKAEMMRTGYAIEYDIVLPHQLRATLETKLISGLFTAGQTNGTSGY 380

Query: 361 EEAAGQGLVAGINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSRAEYRL 420
           EEAAGQGL+AGINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSRAEYRL
Sbjct: 381 EEAAGQGLIAGINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSRAEYRL 440

Query: 421 ILRHDNADMRLTEIGYEIGLVDEERYAIFKKRQMQFENELERLDSIKLKPVSETNKRIQE 480
           ILRHDNADMRLTEIG +IGLVD+ER+  F+ ++ QF+NEL+RL+SIKLKP+ ETN R+Q+
Sbjct: 441 ILRHDNADMRLTEIGRDIGLVDDERWKAFEIKKNQFDNELKRLNSIKLKPIKETNDRVQD 500

Query: 481 LGFKPLTDALTAKEFMRRPQITYAVATDFVGCADEPLDSKVIELLETEIKYEGYIKKALD 540
           LGFKPLTDA+TAKEFMRRP+I YA A  FVG A E LD+K+IELLETEIKYEGYI+KALD
Sbjct: 501 LGFKPLTDAMTAKEFMRRPEIDYATAVSFVGPAAEDLDAKIIELLETEIKYEGYIRKALD 560

Query: 541 QVAKMKRMEEKRIPPHIDWDDIDSIATEARQKFKKINPETLGQASRISGVNPADISILMV 600
           QVAKMKRMEEKRIP +IDWD IDSIATEARQKFKKINPET+GQASRISGVNPADISILM+
Sbjct: 561 QVAKMKRMEEKRIPTNIDWDAIDSIATEARQKFKKINPETIGQASRISGVNPADISILMI 620

Query: 601 YLEGRQKGRK 610
           YLEG  K  +
Sbjct: 621 YLEGNGKAHR 630
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1036

A DNA sequence (GBSx1108) was identified in *S.agalactiae* <SEQ ID 3201> which encodes the amino acid sequence <SEQ ID 3202>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07750 GB:AP001520 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 205/644 (31%), Positives = 362/644 (55%), Gaps = 28/644 (4%)

Query: 35   LLLAIFVALSFVVALLYYQ---------------------KITYELSEVEQIELLNDQTE 73
            ++  + VAL F++AL +YQ                     +I++E    + I  L+ +
Sbjct: 14   VIALLAVALVFLIALSFYQWQLGVIGVLLLLVIAIFSLRARISFERDLEQYISTLSYRVH 73

Query: 74   VSLKSLLEQMPVGVIQFDLETNDIEWFNPYA-ELIFTGDNGHFQSATVKDIITSRRNGTA 132
             + +  + Q+PVG+I ++ +   ++W NPYA E +   +          +++ +   GT
Sbjct: 74   KAGEEAVTQLPVGMILYNDQLR-VQWVNPYAAEHLPKAEIDASLEELSPELVRALEEGTD 132

Query: 133  GQSFEYGDNKYSAYLDTETGVFYFFDNFMGNRRNYDSSMLRPVIGIISIDNYDDIMDTML 192
             Q     +  Y         + YFFD     R +      +PV+  I +DNYD++   M
Sbjct: 133  EQKIVIEEKTYDCTFKPNERLIYFFDITESERMHQQFEESQPVLTFIYLDNYDEVTQGME 192

Query: 193  EADMSKINAFVTSFISDFTQSKNIFYRRVNMDRYYIFTDYSVLNTLIKDKFDILNEFRKR 252
            +   S++ + VTS ++ +    ++F RR   DR+     Y  L  + K KF IL+E R+
Sbjct: 193  DQVRSRLMSQVTSSLNQWANEHDLFLRRTAADRFIAVMSYGSLLAIEKTKFGILDEIRET 252

Query: 253  AQENHLSLTLSMGISYGDGNHNQIGQIALENLNTALVRGGDQIVVRENDSSKKALYFGGG 312
              +  + LTLS+G+ YGD +  ++GQ+A  +L+ AL RGGDQ+ +++     K  ++GG
Sbjct: 253  TGKEKIPLTLSIGVGYGDLSLRELGQLAQSSLDLALGRGGDQVAIKQKTG--KVRFYGGK 310

Query: 313  AVSTIKRSRTRTRAMMTAISDRLKVVDSVFIVGHRKLDMDALGASVGMQFFASNIVNASY 372
            + +  KR+R R R +  A+ D +   D V ++GH+  DMDA+GA++G+   A      ++
Sbjct: 311  SNAMEKRTRVRARVISHALRDFVLESDRVIVMGHKNPDMDAVGAAIGILKIAEVNDREAF 370

Query: 373  VVYDPNDMNSDIERAIDYLQEDGET--RLVSVERAFELITQNSLLVMVDHSKTALTLSKE 430
            VV DPND+N D+ + ++ ++++ +   + ++ E + EL+T+ +LLV+VD  K ++ +
Sbjct: 371  VVLDPNDVNPDVSKLMEEVEKNEQLWDKFITPEESLELMTEETLLVIVDTHKPSMVIEPR 430

Query: 431  FFNKFADVIVVDHHRRDEDFPKNAVLSFIESGASSASELVTELIQFQQAKDKLSRSQASI 490
              +    V+V+DHHRR E+F ++ VL ++E  ASS +ELVTEL+++Q  K K+   +++
Sbjct: 431  LLDYVERVVVLDHHRRGEEFIEDPVLVYMEPYASSTAELVTELLEYQPKKLKMDILESTA 490

Query: 491  LMAGIMLDTRNFASNVTSRTFDVASYLRGLGSNSMAIQKISATDFDEYRLINELILKGER 550
            L+AG+++DT++FA    +RTFD AS+LR  G++++ +QK+   D + Y    +L+   +
Sbjct: 491  LLAGMIVDTKSFAIRTGARTFDAASFLRSHGADTVLVQKLLKEDLNHYVKRAKLVETAKL 550

Query: 551  IYDNIIVATGEEHKVYSHVIASKAADTMLTMAGIEATFVITKNSSN-IGISARSRNNINV 609
              D + +AT  E +  S ++ ++AADT+LTM G+ A+FVI++     + ISARS  ++NV
Sbjct: 551  YRDGMAIATAREEEAVSQLLIAQAADTLLTMKGVVASFVISRRHDGVVSISARSLGDVNV 610

Query: 610  QRIMEKLGGGGHFSFAACQIQDKSVKQVRRMLLEIIDEDLRENS 653
            Q IME L GGGH + AA Q +D ++++    L E ID+ L   S
Sbjct: 611  QLIMESLDGGGHLTNAATQFEDATLEEAEAKLKEAIDQYLEGGS 654
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3203> which encodes the amino acid sequence <SEQ ID 3204>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-18.57   Transmembrane   33 -  49 (   6 -  56)
   INTEGRAL    Likelihood =-10.14   Transmembrane   12 -  28 (   6 -  32)

----- Final Results -----
               bacterial membrane --- Certainty=0.8429(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB07750 GB:AP001520 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 199/659 (30%), Positives = 367/659 (55%), Gaps = 16/659 (2%)

Query: 1   MKKF---RFETIHLI-MMGLILFGLLALCVSIMQSKILILLAIFLVLLFVV-ALLWYQKE 55
           M KF   R+   H+I ++ + L  L+AL     Q  ++ +L + ++ +F + A + ++++
Sbjct: 1   MPKFLLKRWHGYHVIALLAVALVFLIALSFYQWQLGVIGVLLLLVIAIFSLRARISFERD 60

Query: 56  AYQLSDLAHIELLNEQTEDNLKTLLDNMPVGVVQFDQETNAVEWYNPYA-ELIFTTEEGF 114
             Q     +I  L+ +     +  +  +PVG++ ++ +   V+W NPYA E +   E
Sbjct: 61  LEQ-----YISTLSYRVHKAGEEAVTQLPVGMILYNDQLR-VQWVNPYAAEHLPKAEIDA 114

Query: 115 IQNGLIQQIITEKRREDISQTFEVSGNKYTSYIDVSSGIFYFFDSFVGNRQLADASMLRP 174
               L  +++         Q   +    Y      +  + YFFD     R        +P
Sbjct: 115 SLEELSPELVRALEEGTDEQKIVIEEKTYDCTFKPNERLIYFFDITESERMHQQFEESQP 174

Query: 175 VVGIISVDNYDDITDDLSDADTSKINSFVANFIDEFMESKRIFYRRVNMDRYYFFTDFKT 234
           V+  I +DNYD++T  + D   S++ S V + ++++     +F RR   DR+     + +
Sbjct: 175 VLTFIYLDNYDEVTQGMEDQVRSRLMSQVTSSLNQWANEHDLFLRRTAADRFIAVMSYGS 234

Query: 235 LNDLMDNKFSVLEEFRKEAQDAQRPLTLSIGISFGEENHSQIGQVALENLNIALVRGGDQ 294
           L  +   KF +L+E R+      + PLTLSIG+ +G+ +  ++GQ+A  +L++AL RGGDQ
Sbjct: 235 LLAIEKTKFGILDEIRETTGKEKIPLTLSIGVGYGDLSLRELGQLAQSSLDLALGRGGDQ 294

Query: 295 IVIRENADHTNPIYFGGGSVSTVKRSRTRTRAMMTAISDRIKMVDNVFIVGHRKLDMDAL 354
           + I++        ++GG S +  KR+R R R +  A+ D +   D V ++GH+  DMDA+
Sbjct: 295 VAIKQKTGKVR--FYGGKSNAMEKRTRVRARVISHALRDFVLESDRVIVMGHKNPDMDAV 352

Query: 355 GSAVGMQFFAGNIIENSFAVYNPDEMSPDIERAIERLQADGKT--RLISVSQAMGLVTPR 412
           G+A+G+   A      +F V +P++++PD+ + +E ++ + +   + I+  +++ L+T
Sbjct: 353 GAAIGILKIAEVNDREAFVVLDPNDVNPDVSKLMEEVEKNEQLWDKFITPEESLELMTEE 412

Query: 413 SLLVMVDHSKISLTLSKEFYEQFQNVIVVDHHRRDDDFPDNAILTFIESGASSAAELVTE 472
           +LLV+VD  K S+ +     +  + V+V+DHHRR ++F ++ +L ++E  ASS AELVTE
Sbjct: 413 TLLVIVDTHKPSMVIEPRLLDYVERVVVLDHHRRGEEFIEDPVLVYMEPYASSTAELVTE 472

Query: 473 LIQFQNAKKCLNKIQASVLMAGIMLDTKNFSTRVTSRTFDVASYLRSKGSDSVEIQNISA 532
           L+++Q  K  ++ ++++ L+AG+++DTK+F+ R  +RTFD AS+LRS G+D+V +Q +
Sbjct: 473 LLEYQPKKLKMDILESTALLAGMIVDTKSFAIRTGARTFDAASFLRSHGADTVLVQKLLK 532

Query: 533 TDFEEYKQINEIILQGERLGDSIIVAAGEKNHLYSNVIASKAADTILSMAHVEASFVLVE 592
            D   Y +  +++   +   D + +A   +    S ++ ++AADT+L+M  V ASFV+
Sbjct: 533 EDLNHYVKRAKLVETAKLYRDGMAIATAREEEAVSQLLIAQAADTLLTMKGVVASFVISR 592

Query: 593 TASHKIAISARSRSKINVQRVMEKLGGGGHFNLAACQLTDISLPQAKYLLLKTINMTMK 651
                ++ISARS   +NVQ +ME L GGGH   AA Q  D +L +A+  L + I+  ++
Sbjct: 593 RHDGVVSISARSLGDVNVQLIMESLDGGGHLTNAATQFEDATLEEAEAKLKEAIDQYLE 651
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 428/658 (65%), Positives = 547/658 (83%), Gaps = 1/658 (0%)

Query: 1   MKRFRFATVHLVLIGLILFGLLAICVRLFQSYTALLLAIFVALSFVVALLYYQKITYELS 60
           MK+FRF T+HL+++GLILFGLLA+CV + QS   +LLAIF+ L FVVALL+YQK  Y+LS
Sbjct: 1   MKKFRFETIHLIMMGLILFGLLALCVSIMQSKILILLAIFLVLLFVVALLWYQKEAYQLS 60

Query: 61  EVEQIELLNDQTEVSLKSLLEQMPVGVIQFDLETNDIEWFNPYAELIFTGDNGHFQSATV 120
```

```
           ++  IELLN+QTE +LK+LL+ MPVGV+QFD ETN +EW+NPYAELIFT + G  Q+  +
Sbjct: 61  DLAHIELLNEQTEDNLKTLLDNMPVGVVQFDQETNAVEWYNPYAELIFTTEEGFIQNGLI 120

Query: 121 KDIITSRRNGTAGQSFEYGDNKYSAYLDTETGVFYFFDNFMGNRRNYDSSMLRPVIGIIS 180
           + IIT +R     Q+FE   NKY++Y+D  +G+FYFFD+F+GNR+  D+SMLRPV+GIIS
Sbjct: 121 QQIITEKRREDISQTFEVSGNKYTSYIDVSSGIFYFFDSFVGNRQLADASMLRPVVGIIS 180

Query: 181 IDNYDDIMDTMLEADMSKINAFVTSFISDFTQSKNIFYRRVNMDRYYIFTDYSVLNTLIK 240
           +DNYDDI D + +AD SKIN+FV +FI +F +SK IFYRRVNMDRYY FTD+  LN L+
Sbjct: 181 VDNYDDITDDLSDADTSKINSFVANFIDEFMESKRIFYRRVNMDRYYFFTDFKTLNDLMD 240

Query: 241 DKFDILNEFRKRAQENHLSLTLSMGISYGDGNHNQIGQIALENLNTALVRGGDQIVVREN 300
           +KF +L EFRK AQ+    LTLS+GIS+G+ NH+QIGQ+ALENLN ALVRGGDQIV+REN
Sbjct: 241 NKFSVLEEFRKEAQDAQRPLTLSIGISFGEENHSQIGQVALENLNIALVRGGDQIVIREN 300

Query: 301 DSSKKALYFGGGAVSTIKRSRTRTRAMMTAISDRLKVVDSVFIVGHRKLDMDALGASVGM 360
                 +YFGGG+VST+KRSRTRTRAMMTAISDR+K+VD+VFIVGHRKLDMDALG++VGM
Sbjct: 301 ADHTNPIYFGGGSVSTVKRSRTRTRAMMTAISDRIKMVDNVFIVGHRKLDMDALGSAVGM 360

Query: 361 QFFASNIVNASYVVYDPNDMNSDIERAIDYLQEDGETRLVSVERAFELITQNSLLVMVDH 420
           QFFA NI+  S+ VY+P++M+ DIERAI+ LQ DG+TRL+SV +A  L+T  SLLVMVDH
Sbjct: 361 QFFAGNIIENSFAVYNPDEMSPDIERAIERLQADGKTRLISVSQAMGLVTPRSLLVMVDH 420

Query: 421 SKTALTLSKEFFNKFADVIVVDHHRRDEDFPKNAVLSFIESGASSASELVTELIQFQQAK 480
           SK +LTLSKEF+ +F +VIVVDHHRRD+DFP NA+L+FIESGASSA+ELVTELIQFQ AK
Sbjct: 421 SKISLTLSKEFYEQFQNVIVVDHHRRDDDFPDNAILTFIESGASSAAELVTELIQFQNAK 480

Query: 481 DKLSRSQASILMAGIMLDTRNFASNVTSRTFDVASYLRGLGSNSMAIQKISATDFDEYRL 540
             L++ QAS+LMAGIMLDT+NF++ VTSRTFDVASYLR  GS+S+ IQ ISATDF+EY+
Sbjct: 481 KCLNKIQASVLMAGIMLDTKNFSTRVTSRTFDVASYLRSKGSDSVEIQNISATDFEEYKQ 540

Query: 541 INELILKGERIYDNIIVATGEEHKVYSHVIASKAADTMLTMAGIEATFVITKNSSN-IGI 599
           INE+IL+GER+ D+IIVA GE++ +YS+VIASKAADT+L+MA +EA+FV+ + +S+ I I
Sbjct: 541 INEIILQGERLGDSIIVAAGEKNHLYSNVIASKAADTILSMAHVEASFVLVETASHKIAI 600

Query: 600 SARSRNNINVQRIMEKLGGGGHFSFAACQIQDKSVKQVRRMLLEIIDEDLRENSTVEN 657
           SARSR+ INVQR+MEKLGGGGHF+ AACQ+ D S+ Q + +LL+ I+  ++E   VE+
Sbjct: 601 SARSRSKINVQRVMEKLGGGGHFNLAACQLTDISLPQAKYLLLKTINMTMKETGEVES 658
```

A related GBS gene <SEQ ID 8717> and protein <SEQ ID 8718> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:     13.82
GvH: Signal Score (-7.5): -0.890001
     Possible site: 44
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   2.97 threshold:  0.0
   PERIPHERAL  Likelihood =  2.97     574
 modified ALOM score:  -1.09

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
31.3/55.8% over 631aa
  Bacillus subtilis
  EGAD|19304| hypothetical 74.3 kd protein in rpli-cotf intergenic region Insert
characterized
  SP|P37484|YYBT_BACSU HYPOTHETICAL 74.3 KDA PROTEIN IN RPLI-COTF INTERGENIC REGION. Insert
characterized
  GP|467336|dbj|BAA05182.1||D26185 unknown Insert characterized
  GP|2636598|emb|CAB16088.1||Z99124 yybT Insert characterized
```

```
  PIR|S65976|S65976 yybT protein - Insert characterized

ORF00251(364 - 2241 of 2580)
EGAD|19304|BS4045(20 - 651 of 659) hypothetical 74.3 kd protein in rpli-cotf intergenic
region {Bacillus subtilis}SP|P37484|YYBT_BAC
SU HYPOTHETICAL 74.3 KDA PROTEIN IN RPLI-COTF INTERGENIC
REGION.GP|467336|dbj|BAA05182.1||D26185 unknown {Bacillus subtilis}GP|26365
98|emb|CAB16088.1||Z99124 yybT {Bacillus subtilis}PIR|S65976|S65976 yybT protein - Bacillus
subtilis
%Match = 18.5
%Identity = 31.2  %Similarity = 55.8
Matches = 197  Mismatches = 271  Conservative Sub.s = 155

258       288       318       348       378       408       438       468
N***CSPLFIRGVLCYN*VLRGYLMKRFRFATVHLVLIGLILFGLLAICVRLFQSYTALLLAIFVALSFVVALLYYQKIT
                             |    |    :  :  |:|   |   |   :|     | : | |:: ::
                          MPSFYEKPLFRYPIYALIALSIITILISFYFNWILGTVEVLLLAVILFFIKRAD
                                  10        20        30        40        50

522       552       582       612                 666       696
YEL-SEVEQ-IELLNDQTEVSLKSLLEQMPVGVIQFDLETNDIEWFNPYAELIFTGDN--GHFQSATVKDIITSRRNGTA
  :  |::  |  |: : :   :  | :||:|:: |: :   ||| ||:    |      |     | : ::   :
SLIRQEIDAYISTLSYRLKKVGEEALMEMPIGIMLFN-DQYYIEWANPFLSSCFNESTLVGRSLYDTCESVVPLIKQEVE
              70        80        90        100       110       120       130

726       756       786       816       846       876       906       936
GQSFEYGDNKYSAYLDTETGVFYFFDNFMGNRRNYDSSMLRPVIGIISIDNYDDIMDTMLEADMSKINAXVTSFXSDFTQ
 ::    | |:   :  :  ::||||     :        | |:  | :|||||:   : :   | :|: |||: : : |
SETVTLNDRKFRVVIKRDERLLYFFDVTEQIQIEKLYENERTVLAYIFLDNYDDVTQGLDDQTRSTMNSQVTSLLNAWAQ
              150       160       170       180       190       200       210

966       996      1026      1056      1086      1116      1146      1176
SKNIFYRRVNMDRYYIFTDYSVLNTLIKDKFDILNEFRKRAQENHLSLTLSMGISYGDGNHNQIGQIALENLNTALVRGG
   || :| : :|:    :  :|  |   || ||:| |::   : ::||||:|:     :  ::| :|  :|: || |||
EYGIFLKRTSSERFIAVLNEHILTELENSKFSILDEVREKTSFDGVALTLSVGVGASVSSLKELGDLAQSSLDLALGRGG
              230       240       250       260       270       280       290

1206      1236      1266      1296      1326      1356      1386      1416
DQIVVRENDSSKKALYFGGGAVSTIKRSRTRTRAMMTAISDRLKVVDSVFIVGHRKLDMDALGASVGMQFFASNIVNASY
||: ::  :   |  ::||      ||:| | | :  |: : :    :| |:||:  |||::||::|:   |       :
DQVAIKLPNGKVK--FYGGKTNPMEKRTRVRARVISHALKEIVTESSNVIIMGHKFPDMDSIGAAIGILKVAQANNKDGF
                310       320       330       340       350       360       370

1446      1476      1500      1530      1560      1590      1620      1650
VVYDPNDMNSDIERAIDYLQEDGE--TRLVSVERAFELITQNSLLVMVDHSKTALTLSKEFFNKFADVIVVDHHRRDEDF
:| ||| : | ::| |  :::  |  :|::: | | |:   ::|||:||  | :| : : : ||   ::|:||||| |:|
IVIDPNQIGSSVQRLIGEIKKYEELWSRFITPEEAMEISNDDTLLVIVDTHKPSLVMEERLVNKIEHIVVIDHHRRGEEF
                390       400       410       420       430       440       450

1680      1710      1740      1770      1800      1830      1860      1890
PKNAVLSFIESGASSASELVTELIQFQQAKDKLSRSQASILMAGIMLDTRNFASNVTSRTFDVASYLRGLGSNSMAIQKI
 :: :| ::|  ||| :||||||:::|  : |::  :|: |:|||::||::|:    ||||| |||||  |:::: :||
IRDPLLVYMEPYASSTAELVTELLEYQPKRLKINMIEATALLAGIIVDTKSFSLRTGSRTFDAASYLRAKGADTVLVQKF
                470       480       490       500       510       520       530

1920      1950                2004      2034      2064      2091      2121
SATDFDEYRLINELILKGERIYDNIIVAT--GEEHKVYSHVIASKAADTMLTMAGIEATFVITK-NSSNIGISARSRNNI
     | |    :||       ||| :|:    | : : :|: ::|||::|:|: :||:| : : :   : |||||   :
LKETVDSYIKRAKLIQHTVLYKDNIAIASLPENEEEYFDQVLIAQAADSLLSMSEVEASFAVARRDEQTVCISARSLGEV
                550       560       570       580       590       600       610

2151      2181      2211      2241      2271      2301      2331      2361
NVQRIMEKLGGGGHFSFAACQIQDKSVKQVRRMLLEIIDEDLRENSTVENRRD*LR*KLFFYKMLRGKEKKVRLRKYLLV
||| ||| | ||||:: || |:   || :    |   ||| :
NVQIIMEALEGGGHLTNAATQLSGISVSEALERLKHAIDEYFEGGVQR
                630       640       650
```

SEQ ID 8718 (GBS10) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 6; MW 98kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 7; MW 73kDa).

The GST-fusion protein was purified as shown in Figure 189, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1037

A DNA sequence (GBSx1109) was identified in *S.agalactiae* <SEQ ID 3205> which encodes the amino acid sequence <SEQ ID 3206>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4643(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA43972 GB:X62002 ribosomal protein L9 [Bacillus
           stearothermophilus]
 Identities = 80/149 (53%), Positives = 105/149 (69%), Gaps = 2/149 (1%)

Query: 1   MKVIFLQDVKGKGKKGEVKEVPTGYAQNFLLKKNLAKEATTQAIGELKGKQKSEEKAQAE 60
           MKVIFL+DVKGKGKKGE+K V  GYA NFL K+ LA EAT   +  L+ +++ E++  AE
Sbjct: 1   MKVIFLKDVKGKGKKGEIKNVADGYANNFLFKQGLAIEATPANLKALEAQKQKEQRQAAE 60

Query: 61  ILAQAKELKTQLESETTRVQFIEKVGPDGRTFGSITAKKIAEELQKQYGIKIDKRHIDLD 120
            LA AK+LK QLE T  +    K G  GR FGSIT+K+IAE LQ Q+G+K+DKR I+L
Sbjct: 61  ELANAKKLKEQLEKLTVTIP--AKAGEGGRLFGSITSKQIAESLQAQHGLKLDKRKIELA 118

Query: 121 HTIRAIGKVEVPVKLHKQVSSQIKLDIKE 149
              IRA+G   VPVKLH +V++ +K+ + E
Sbjct: 119 DAIRALGYTNVPVKLHPEVTATLKVHVTE 147
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3207> which encodes the amino acid sequence <SEQ ID 3208>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4630(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 119/150 (79%), Positives = 138/150 (91%)

Query: 1   MKVIFLQDVKGKGKKGEVKEVPTGYAQNFLLKKNLAKEATTQAIGELKGKQKSEEKAQAE 60
           MKVIFL DVKGKGKKGE+KEVPTGYAQNFL+KKNLAKEAT+Q+IGELKGKQK+EEKAQAE
Sbjct: 1   MKVIFLADVKGKGKKGEIKEVPTGYAQNFLIKKNLAKEATSQSIGELKGKQKAEEKAQAE 60

Query: 61  ILAQAKELKTQLESETTRVQFIEKVGPDGRTFGSITAKKIAEELQKQYGIKIDKRHIDLD 120
           ILA+A+ +K  L+ + TRVQF EKVGPDGRTFGSITAKKI+EELQKQ+G+K+DKRHI LD
Sbjct: 61  ILAEAQAVKAVLDEDKTRVQFQEKVGPDGRTFGSITAKKISEELQKQFGVKVDKRHIVLD 120
```

```
Query: 121 HTIRAIGKVEVPVKLHKQVSSQIKLDIKEA 150
           H IRAIG +EVPVKLHK+V+++IKL I EA
Sbjct: 121 HPIRAIGLIEVPVKLHKEVTAEIKLAITEA 150
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1038

A DNA sequence (GBSx1110) was identified in *S.agalactiae* <SEQ ID 3209> which encodes the amino acid sequence <SEQ ID 3210>. This protein is predicted to be DNA polymerase III delta prime subunit (dnaB). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.43    Transmembrane  204 - 220 ( 204 - 220)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2423> which encodes the amino acid sequence <SEQ ID 2424>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27    Transmembrane  210 - 226 ( 210 - 226)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 397/450 (88%), Positives = 431/450 (95%), Gaps = 1/450 (0%)

Query: 3   EVSELRVQPQDLLAEQAVLGSIFISPEKLIMVREFISPDDFYKYSHKVIFRAMITLADRN 62
           EV+ELRVQPQDLLAEQ+VLGSIFISP+KLI VREFISPDDFYKY+HK+IFRAMITL+DRN
Sbjct: 8   EVAELRVQPQDLLAEQSVLGSIFISPDKLIAVREFISPDDFYKYAHKIIFRAMITLSDRN 67

Query: 63  DAIDAATVRNILDDQGDLQNIGGLGYIVELVNSVPTSANAEFYAKIVSEKAMLRDIISKL 122
           DAIDA T+R ILDDQ DLQ+IGGL YIVELVNSVPTSANAE+YAKIV+EKAMLRDII++L
Sbjct: 68  DAIDATTIRTILDDQDDLQSIGGLSYIVELVNSVPTSANAEYYAKIVAEKAMLRDIIARL 127

Query: 123 TDTVNMAY-EGNDSDEIIATAEKALVDINEHSNRSGFRKISDVLKVNYENLELRSQQTSD 181
           T++VN+AY E    +E+IA  E+AL+++NEHSNRSGFRKISDVLKVNYE LE RS+QTS+
Sbjct: 128 TESVNLAYDEILKPEEVIAGVERALIELNEHSNRSGFRKISDVLKVNYEALEARSKQTSN 187

Query: 182 VTGLPTGFRDLDRITTGLHPDQLIILAARPAVGKTAFVLNIAQNVGTKQNRPVAIFSLEM 241
           VTGLPTGFRDLD+ITTGLHPDQL+ILAARPAVGKTAFVLNIAQNVGTKQ + VAIFSLEM
Sbjct: 188 VTGLPTGFRDLDKITTGLHPDQLVILAARPAVGKTAFVLNIAQNVGTKQKKTVAIFSLEM 247

Query: 242 GAESLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALADAPIYIDDTPGIKITEIR 301
           GAESLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALA+APIYIDDTPGIKITEIR
Sbjct: 248 GAESLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALAEAPIYIDDTPGIKITEIR 307

Query: 302 ARSRKLSQEVDDGLGLIVIDYLQLISGTRPENRQQEVSEISRQLKILAKELKVPVIALSQ 361
           ARSRKLSQEVD GLGLIVIDYLQLI+GT+PENRQQEVS+ISRQLKILAKELKVPVIALSQ
Sbjct: 308 ARSRKLSQEVDGGLGLIVIDYLQLITGTKPENRQQEVSDISRQLKILAKELKVPVIALSQ 367

Query: 362 LSRGVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRREGEEAEEIVEDNTVEVIL 421
           LSRGVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYR+E ++AEE VEDNT+EVIL
```

```
Sbjct: 368 LSRGVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRKECDDAEEAVEDNTIEVIL 427

Query: 422 EKNRAGARGTVKLMFQKEYNKFSSIAQFEE 451
           EKNRAGARGTVKLMFQKEYNKFSSIAQFEE
Sbjct: 428 EKNRAGARGTVKLMFQKEYNKFSSIAQFEE 457
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1039

A DNA sequence (GBSx1111) was identified in *S.agalactiae* <SEQ ID 3211> which encodes the amino acid sequence <SEQ ID 3212>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4909(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3213> which encodes the amino acid sequence <SEQ ID 3214>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3467(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 77/90 (85%), Positives = 84/90 (92%)

Query: 1    MSDAFADVAKMKKIKEDIKSHEGQMVELTLENGRKREKNKIGRLIEVYPSLFIVEYKDTA 60
            MSDAF DVAKMKKIKEDI++HEGQ+VELTLENGRKREKNKIGRLIEVY SLFI+EY D++
Sbjct: 11   MSDAFTDVAKMKKIKEDIRAHEGQLVELTLENGRKREKNKIGRLIEVYSSLFIIEYSDSS 70

Query: 61   AVPGAIDNTYVESYTYSDILTEKTLIRYFD 90
              PGAIDN+YVESYTYSDILTEKTLIRY D
Sbjct: 71   DTPGAIDNSYVESYTYSDILTEKTLIRYLD 100
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1040

A DNA sequence (GBSx1112) was identified in *S.agalactiae* <SEQ ID 3215> which encodes the amino acid sequence <SEQ ID 3216>. This protein is predicted to be 30S ribosomal protein S4 (rpsD). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2937(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00397 GB:AF008220 ribosomal protein S4 [Bacillus subtilis]
 Identities = 138/201 (68%), Positives = 158/201 (77%), Gaps = 1/201 (0%)

Query: 1   MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS 60
           M+RYTGPSWK SRRLG+SL+GTGKEL +R Y PG HGP  R KLSEYGLQL EKQKLR
Sbjct: 1   MARYTGPSWKLSRRLGISLSGTGKELEKRPYAPGPHGPGQRKKLSEYGLQLQEKQKLRHM 60

Query: 61  YGLGEKQFRNLFVQATKAKEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI 120
           YG+ E+QFR LF +A K   G  G NFM+LL+ RLDNVVY+LGLA TRRQARQ VNHGHI
Sbjct: 61  YGVNERQFRTLFDKAGKLA-GKHGENFMILLDSRLDNVVYKLGLARTRRQARQLVNHGHI 119

Query: 121 LVDGKRVDIPSYRVTPGQVISVREKSMKVPAILEAVEATLGRPAFVSFDAEKLEGSLTRL 180
           LVDG RVDIPSY V PGQ I VREKS  +  I E+VE      P +++FDAEKLEG+ TRL
Sbjct: 120 LVDGSRVDIPSYLVKPGQTIGVREKSRNLSIIKESVEVNNFVPEYLTFDAEKLEGTFTRL 179

Query: 181 PERDEINPEINEALVVEFYNK 201
           PER E+ PEINEAL+VEFY++
Sbjct: 180 PERSELAPEINEALIVEFYSR 200
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3217> which encodes the amino acid sequence <SEQ ID 3218>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2937(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/203 (99%), Positives = 201/203 (99%)

Query: 1   MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS 60
           MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS
Sbjct: 1   MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS 60

Query: 61  YGLGEKQFRNLFVQATKAKEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI 120
           YGLGEKQFRNLFVQATK KEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI
Sbjct: 61  YGLGEKQFRNLFVQATKIKEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI 120

Query: 121 LVDGKRVDIPSYRVTPGQVISVREKSMKVPAILEAVEATLGRPAFVSFDAEKLEGSLTRL 180
           LVDGKRVDIPSYRV PGQVISVREKSMKVPAILEAVEATLGRPAFVSFDAEKLEGSLTRL
Sbjct: 121 LVDGKRVDIPSYRVDPGQVISVREKSMKVPAILEAVEATLGRPAFVSFDAEKLEGSLTRL 180

Query: 181 PERDEINPEINEALVVEFYNKML 203
           PERDEINPEINEALVVEFYNKML
Sbjct: 181 PERDEINPEINEALVVEFYNKML 203
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1041

A DNA sequence (GBSx1113) was identified in *S.agalactiae* <SEQ ID 3219> which encodes the amino acid sequence <SEQ ID 3220>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4067(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF98302 GB:AF243383 unknown; Orf3 [Lactococcus lactis subsp.
          lactis]
 Identities = 46/97 (47%), Positives = 69/97 (70%)

Query: 1   MNLNDRLKIEEMEEKYDSFKPRINALVEAIDDFQKHYEDYVKLREFYGSEDWFRLSEQTE 60
            M+  D   I++ME KYD+F P +  L+++++ F   Y +Y++LR FYGSE WF   E  +
Sbjct: 1   MDNKDIELIQQMENKYDTFMPVLTNLIDSVEKFNSIYNNYIELRNFYGSEKWFEYMEIEK 60

Query: 61 NNLKCGVLSEDQLFDFIGEHNELVGQFLDMSSQMYRH 97
              +KCGVL+EDQLFD I +HNEL+G  LD++S+MY++
Sbjct: 61 IPVKCGVLTEDQLFDMISDHNELLGVLLDLTSKMYKN 97
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3221> which encodes the amino acid sequence <SEQ ID 3222>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3465(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 48/98 (48%), Positives = 74/98 (74%)

Query: 1   MNLNDRLKIEEMEEKYDSFKPRINALVEAIDDFQKHYEDYVKLREFYGSEDWFRLSEQTE 60
            M   D+L +E+ME+ Y++F P++  L+EA+D F++HYE+Y  LR FY S++WFRL+ Q
Sbjct: 1   MTKQDQLIVEKMEQTYEAFSPKLANLIEALDAFKEHYEEYATLRNFYSSDEWFRLANQPW 60

Query: 61 NNLKCGVLSEDQLFDFIGEHNELVGQFLDMSSQMYRHL 98
           +++ CGVLSED LFD IG+HN+L+   LD++  MY+H+
Sbjct: 61 DDIPCGVLSEDLLFDMIGDHNQLLADILDLAPIMYKHM 98
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1042

A DNA sequence (GBSx1114) was identified in *S.agalactiae* <SEQ ID 3223> which encodes the amino acid sequence <SEQ ID 3224>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04438 GB:AP001509 transcriptional regulator (TetR/AcrR
           family) [Bacillus halodurans]
 Identities = 47/181 (25%), Positives = 95/181 (51%), Gaps = 16/181 (8%)

Query: 4    DTRREKTKRAIEAAMITLLKDQSFDEISTINLTKTAGISRSSFYTHYKDKYEMIDQYQQS 63
```

```
           D R++ T+  ++ +++ L++++    I+   +   A I+RS+FY+HY D Y+++ Q +
Sbjct: 6   DRRKKYTRMLLKESLMKLMQEKPLSNITIKEICDLADINRSTFYSHYTDLYDLLYQIEDE 65

Query: 64  LFNKV-EYIFDRNQFKKEDAL-----LEIFQFLDRESLFAALLTQNGTKEIQTYILNKLQ 117
           +   + E +   N  K E+AL     L ++   +RES    L ++ G    Q     K
Sbjct: 66  IIKDLSEALSSYNYTKDEEALQMTENLLVYIANNRESC-QTLFSEYGDPSFQ-----KKV 119

Query: 118 LMLSKELPVVNP---DATKSDINRLYYSVYLSHAIFGVYQMWITRGKKESPQQITQVLLSL 175
           +ML+ +  +  P     TK DI+  Y S+Y+ +    + Q W+  G K+SP+++ ++++ L
Sbjct: 120 MMLAHDHVIKTPLVGKHTKPDISE-YVSLYIVNGSIHIVQSWLKNGLKQSPKEMAELIIKL 179
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3225> which encodes the amino acid sequence <SEQ ID 3226>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04438 GB:AP001509 transcriptional regulator (TetR/AcrR
           family) [Bacillus halodurans]
 Identities = 47/180 (26%), Positives = 88/180 (48%), Gaps = 18/180 (10%)

Query: 4   RKENTKQAILKAMVMLLKTESFDDITTVKLSKRAGISRSSFYTHYKDKYEMIDYYQQTFF 63
           RK+ T+  + ++++ L++ +   +IT  ++   A I+RS+FY+HY D Y+++   +
Sbjct: 8   RKKYTRMLLKESLMKLMQEKPLSNITIKEICDLADINRSTFYSHYTDLYDLLYQIEDEII 67

Query: 64  HKLEYIFEKKYQNKEQAFLEVFEFL-----QREQLLSSLLSANGTKEIQAFIINKVRLL- 117
             L          K++  L++ E L        +   +L S  G    Q     KV +L
Sbjct: 68  KDLSEALSSYNYTKDEEALQMTENLLVYIANNRESCQTLFSEYGDPSFQ----KKVMMLA 123

Query: 118 ----ITTDLQDKFSTEELSQTEKEYQSIYLAHAFFGVCQSWIAKGKKESPQEMTQFVLKM 173
               I T L  K +  ++S    EY S+Y+ +    + QSW+  G K+SP+EM + ++K+
Sbjct: 124 HDHVIKTPLVGKHTKPDIS----EYVSLYIVNGSIHIVQSWLKNGLKQSPKEMAELIIKL 179
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 100/179 (55%), Positives = 134/179 (73%), Gaps = 2/179 (1%)

Query: 1   MVNDTRREKTKRAIEAAMITLLKDQSFDEISTINLTKTAGISRSSFYTHYKDKYEMIDQY 60
           MVN  R+E TK+AI  AM+ LLK +SFD+I+T+ L+K AGISRSSFYTHYKDKYEMID Y
Sbjct: 1   MVN--RKENTKQAILKAMVMLLKTESFDDITTVKLSKRAGISRSSFYTHYKDKYEMIDYY 58

Query: 61  QQSLFNKVEYIFDRNQFKKEDALLEIFQFLDRESLFAALLTQNGTKEIQTYILNKLQLML 120
           QQ+ F+K+EYIF++    KE A LE+F+FL RE L ++LL+ NGTKEIQ +I+NK++L++
Sbjct: 59  QQTFFHKLEYIFEKKYQNKEQAFLEVFEFLQREQLLSSLLSANGTKEIQAFIINKVRLLI 118

Query: 121 SKELPVVNPDATKSDINRLYYSVYLSHAIFGVYQMWITRGKKESPQQITQVLLSLLPQT 179
           + +L         S   + Y S+YL+HA FGV Q WI +GKKESPQ++TQ +L +L  T
Sbjct: 119 TTDLQDKFSTEELSQTEKEYQSIYLAHAFFGVCQSWIAKGKKESPQEMTQFVLKMLTST 177
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1043

A DNA sequence (GBSx1115) was identified in *S.agalactiae* <SEQ ID 3227> which encodes the amino acid sequence <SEQ ID 3228>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.35   Transmembrane  790 - 806 ( 787 - 808)
   INTEGRAL    Likelihood = -7.32   Transmembrane  707 - 723 ( 703 - 725)
   INTEGRAL    Likelihood = -7.11   Transmembrane  637 - 653 ( 630 - 659)
   INTEGRAL    Likelihood = -6.32   Transmembrane  678 - 694 ( 672 - 698)
   INTEGRAL    Likelihood = -1.44   Transmembrane   55 -  71 (  55 -  73)
   INTEGRAL    Likelihood = -0.22   Transmembrane  732 - 748 ( 730 - 748)

----- Final Results -----
               bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10287> which encodes amino acid sequence <SEQ ID 10288> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12856 GB:Z99109 alternate gene name: yixE~similar to phage
           infection protein [Bacillus subtilis]
 Identities = 227/783 (28%), Positives = 387/783 (48%), Gaps = 60/783 (7%)

Query: 45  KAIIKSPKLWITMAGVALIPTLYNVIFLSSMWDPYGNTKNLPVAVVNQDKSAKLNGKTIS 104
           K I+ S KL I +  +  +P +Y+ +FL + WDPYG    LPV VVNQDK A   G+ +
Sbjct: 9   KDIVTSKKLLIPIIAILFVPLIYSGVFLKAYWDPYGTVDQLPVVVVNQDKGATYEGEKLQ 68

Query: 105 IGKDMEDNLSKNDSLDFHFTT-AKRAEKELEKGHYYMVITFPKDLSRKATTLMTEKPERL 163
           IG D+   L  N++ D+HF+    ++ K+L    YY+V+  P+D S+ A+T++ + P++L
Sbjct: 69  IGDDLVKELKDNNNFDWHFSNDLDQSLKDLLNQKYYLVVEIPEDFSKNASTVLDKNPKKL 128

Query: 164 NITYKTTKGRSFVASKMSETAANKLKDEVAESITGTYTESVFKNMGSMKTGINKAADGSQ 223
           ++ Y T  G ++V + + E A +KLK  V++ +T  YT+ +F N   +  G++ A+ G++
Sbjct: 129 DLKYHTNAGSNYVGATIGEKAIDKLKASVSKEVTEQYTKVIFDNFKDIAKGLSDASSGAK 188

Query: 224 ELLNGSNKLQDGSQTLTSNLDVLASSSQTFSGGANKLNSGINLYTDGVGTLSNGLETLSD 283
           ++ +G+   ++GS  L  NL  L  S+ T S    +L  G    T G+ +L + L    D
Sbjct: 189 KIDDGTKDAKNGSAQLKENLAKLKESTATISDKTAQLADGAAQVTSGIQSLDSSLGKFQD 248

Query: 284 GVTAYTTGVHKLSEGSQKLDDKSQALV-------EGSEKLTDGLQQLSQATQLKPEQERT 336
                     +L+ GS +L  K   L+       +G+  LT+GL QL+   Q   E+
Sbjct: 249 SSNQIYDKSSQLAAGSGELTSKMNELLAGLQNVQKGTPNLTNGLDQLNSKVQEGSEKAAK 308

Query: 337 LQNLSDG--LKNLNQIITNLQSTATTDSDTNSKLFNFLSTIESSTKALMNTAAADKQKQM 394
            + + +   L  L   + NL+ + T   +   +L +F +++++  +A  N    +    +
Sbjct: 309 AEKIINALDLTKLETAVNNLEKSETAMKEFKKQLTDFENSLKNRDQAFKN--VINSSDFL 366

Query: 395 TAVQST----SAFKSLTPEQQSQITSAVTGTPTSAE-TIAANISSNIENMKTVLSEASSS 449
           TA Q +    S  K L           ++  PT+ +    A I S++E++K  +++  +
Sbjct: 367 TAEQKSQLINSVEKKLPQVDAPDFDQILSQLPTADQLPDIATIKSSLEDVKAQVAQVKAM 426

Query: 450 APSN----NGSQNLQTLSGTANNLVLKAISDLDKIQKLPTATKQLYQGSQTLTKGITDYT 505
             +     NG++ +Q                 D I +L     ++Y GSQ LT G T  T
Sbjct: 427 PEATSKLYNGAKTIQ-----------------DAIDRLTEGADKIYNGSQKLTDGQTKLT 469

Query: 506 NAVGQLRKGAVTLDSKSNQLISGTQKASQGAQTLDSKSDQLRDGAGQLASGSDRIADGSN 565
             +G+  K      + S QL++G              S Q+  G  +L  GS ++  GS+
Sbjct: 470 AGIGEYNKQFAKAKAGSEQLVTG--------------SSQVSGGLFKLLDGSKQVQSGSS 515

Query: 566 KLAGGGHQLTDGLTELSGGVSQLSSSLGKAGDQLSMVSVNKDNANAVSSPVTIKHEDYDS 625
           KLA G   L  GL +L  G  +LSS L  A DQ   +  +       + PV  K +   S
Sbjct: 516 KLADGSASLDTGLGKLLDGTGELSSKLKDAADQTGDIDADDQTYGMFADPVKTKDDAIHS 575

Query: 626 VDTNGVGMAPYMISVALMVVALSANVIFAKALSGKEPANRFSWAKNK---LLINGFIATL 682
           V   G G+ PY++S+ L V  +   V+F    +   P N F W  +K   +++ G I +L
Sbjct: 576 VPNYGTGLTPYILSMGLYVGGIMLTVVFPLKEASGRPRNGFEWFFSKFNVMMLVGIIQSL 635

Query: 683 -AATILFFAVQFIGLKPDYPGKTYFIILLTAWTLMALVTALVGWDNRYGSFLSLLILLFQ 741
```

```
             AT+L      IGL+ +   + Y   ++T+   +A++  L       G F++++IL+ Q
Sbjct: 636 IVATVLLLG---IGLEVESTWRFYVFTIITSLAFLAIIQFLATTMGNPGRFIAVIILVLQ 692

Query: 742 LGSSAGTYPIELSPKFFQTIQPFLPMTYSVSGLRETISLTGDVNHQWRMLVIFLVSSMIL 801
           LG+S GT+P+EL P F+Q I   LPMTYS++G R  IS  GD  + W+M  + +  ++++
Sbjct: 693 LGASGGTFPLELLPNFYQVIHGALPMTYSINGFRAVIS-NGDFGYMWQMAGVLIGIALVM 751

Query: 802 ALL 804
             L
Sbjct: 752 IAL 754
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2017> which encodes the amino acid sequence <SEQ ID 2018>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.29    Transmembrane  735 - 751 ( 729 - 754)
   INTEGRAL    Likelihood = -5.79    Transmembrane  582 - 598 ( 580 - 601)
   INTEGRAL    Likelihood = -3.66    Transmembrane  652 - 668 ( 650 - 669)
   INTEGRAL    Likelihood = -2.97    Transmembrane   14 -  30 (  14 -  34)
   INTEGRAL    Likelihood = -2.66    Transmembrane  623 - 639 ( 622 - 641)

----- Final Results -----
               bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 360/779 (46%), Positives = 508/779 (64%), Gaps = 32/779 (4%)

Query: 40  MLDELKAIIKSPKLWITMAGVALIPTLYNVIFLSSMWDPYGNTKNLPVAVVNQDKSAKLN 99
           ML+ELK +IK+PKL ITM GVAL+P LYN+ FL SMWDPYG   +LP+AVVN DK AK
Sbjct: 1   MLEELKTLIKNPKLMITMIGVALVPALYNLSFLGSMWDPYGRVNDLPIAVVNHDKPAKRA 60

Query: 100 GKTISIGKDMEDNLSKNDSLDFHFTTAKRAEKELEKGHYYMVITFPKDLSRKATTLMTEK 159
            K+++IG DM D +SK+  L++HF +AK+A++ L++G YYMVIT P+DLS++A TL+  +
Sbjct: 61  DKSLTIGNDMVDKMSKSKDLEYHFVSAKQAQEGLKEGDYYMVITLPEDLSQRAATLLNPE 120

Query: 160 PERLNITYKTTKGRSFVASKMSETAANKLKDEVAESITGTYTESVFKNMGSMKTGINKAA 219
           P++L I Y+T+KG   VA+KM ETA  KLK+ V+++IT TYT +VF +M  +++G+ +A+
Sbjct: 121 PQKLTIRYQTSKGHGMVAAKMGETAMAKLKESVSQNITKTYTSAVFSSMTDLQSGLKEAS 180

Query: 220 DGSQELLNGSNKLQDGSQTLTSNLDVLASSSQTFSGGANKLNSGINLYTDGVGTLSNGLE 279
            GSQ L +G+   Q GSQTL++NL  L  +SQ F  G  +L SG+  YTDGV  + NGL
Sbjct: 181 AGSQALASGAKTAQAGSQTLSTNLAALTGASQQFQQGTGRLTSGLTTYTDGVNQVKNGLG 240

Query: 280 TLSDGVTAYTTGVHKLSEGSQKLDDKSQALVEGSEKLTDGLQQLSQATQLKPEQERTLQN 339
           TLS  +  Y  GV +LS+G+ +L+               GL QL+QAT L  E+ + +Q+
Sbjct: 241 TLSTDIPNYLNGVSRLSQGASQLNQ--------------GLSQLTQATTLSDEKAKGIQS 286

Query: 340 LSDGLKNLNQIITNLQSTATTDSDTN---SKLFNFLSTIESSTKALMNTAAADKQKQMTA 396
           L  GL  LNQ I  L +  +T    N    +L N L  I  + K ++    A + ++++A
Sbjct: 287 LIVGLPVLNQGIQQLNTELSTLQPPNLNADELGNSLGAIAQAAKQVIAEETAAQNEELSA 346

Query: 397 VQSTSAFKSLTPEQQSQITSAVTGTPTSAETIAAN-ISSNIENMKTVLSEASSSAPSNNG 455
           +Q+TS ++SLT EQQ ++ +A++ +  S    AA  I S+++ + T L   S    S
Sbjct: 347 LQATSVYQSLTAEQQGELAAALSQSDKSQTVSAAQTILSSVQTLSTSLQSLSQEDQSKQL 406

Query: 456 SQNLQTLSGTANNLVLKAISDLDKIQKLPTATKQLYQGSQTLTKGITDYTNAV----GQL 511
            Q  + ++ AN             Q LP A+  L + S  L K        V     QL
Sbjct: 407 EQLKEAVAQIANQ----------SNQALPGASSALTELSTGLAKVNGSLNQQVLPGSNQL 456

Query: 512 RKGAVTLDSKSNQLISGTQKASQGAQTLDSKSDQLRDGAGQLASGSDRIADGSNKLAGGG 571
             G   L+  +  + SG  K S+GA  L SKS +L DG+ QL+ G+ ++ADGS++L+ GG
Sbjct: 457 TTGLAQLNRYNTAIGSGVIKLSEGANALSSKSGELLDGSHQLSEGATKLADGSSQLSQGG 516
```

```
Query: 572 HQLTDGLTELSGGVSQLSSSLGKAGDQLSMVSVNKDNANAVSSPVTIKHEDYDSVDTNGV 631
           HQLT GLTELS G+S L+ SL KA  QLS+VSV   NA AV+ P+ +  +D D V TNG+
Sbjct: 517 HQLTSGLTELSTGLSTLNGSLAKASQQLSLVSVTDKNAKAVAKPLVLNEKDKDGVKTNGI 576

Query: 632 GMAPYMISVALMVVALSANVIFAKALSGKEPANRFSWAKNKLLINGFIATLAATILFFAV 691
           GMAPYMI+V+LMVVALS NVIFA +LSG+    +++ WAK K +INGFI+T+ + +L+ A+
Sbjct: 577 GMAPYMIAVSLMVVALSTNVIFANSLSGRPVKDKWDWAKQKFVINGFISTMGSIVLYLAI 636

Query: 692 QFIGLKPDYPGKTYFIILLTAWTLMALVTALVGWDNRYGSFLSLLILLFQLGSSAGTYPI 751
           Q +G +  Y  +T  I+L+ WT MALVTALVGWD+RYGSF SL++LL Q+GSS G+YPI
Sbjct: 637 QLLGFEARYGMETLGFIMLSGWTFMALVTALVGWDDRYGSFASLVMLLLQVGSSGGSYPI 696

Query: 752 ELSPKFFQTIQPFLPMTYSVSGLRETISLTGDVNHQWRMLVIFLVSSMILALLIYRKQE 810
           ELS  FFQ + PFLPMTY VSGLR+TISL+G +  + ++L  FL++ M+LALLIYR ++
Sbjct: 697 ELSGAFFQKLHPFLPMTYVVSGLRQTISLSGHIGVEVKVLTGFLLAFMVLALLIYRPKK 755
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1044

A DNA sequence (GBSx1116) was identified in *S.agalactiae* <SEQ ID 3229> which encodes the amino acid sequence <SEQ ID 3230>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2664(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1045

A DNA sequence (GBSx1117) was identified in *S.agalactiae* <SEQ ID 3231> which encodes the amino acid sequence <SEQ ID 3232>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.45   Transmembrane   48 -  64 (  45 -  69)
   INTEGRAL    Likelihood = -1.49   Transmembrane   71 -  87 (  71 -  87)

----- Final Results -----
               bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9441> which encodes amino acid sequence <SEQ ID 9442> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25222 GB:M87483 ORF 1 [Lactococcus lactis]
 Identities = 50/88 (56%), Positives = 66/88 (74%), Gaps = 1/88 (1%)

Query: 2   TGKIFSMSKEELSYLPVIKLFKNQGVYNGLIGLFLLYGLYISQNQ-EIVAVFLINVLLVA 60
```

```
            T ++F+M KEEL     V  LFKNQG+YNGLIGL L+Y ++ S  Q EIV + LI ++LVA
Sbjct: 32  TSRVFNMGKEELERSSVQTLFKNQGIYNGLIGLGLIYAIFFSSAQLEIVRLLLIYIILVA 91

Query: 61  IYGALTVDKKILLKQGGLPILALLTFLF 88
            +YG+LT +KKI+L QGGL ILAL++  F
Sbjct: 92  LYGSLTSNKKIILTQGGLAILALISSFF 119
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8719> and protein <SEQ ID 8720> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      4.19
GvH: Signal Score (-7.5): -3.99
     Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value:  -9.45 threshold:  0.0
   INTEGRAL    Likelihood = -9.45   Transmembrane   87 - 103 (  84 - 108)
   INTEGRAL    Likelihood = -1.49   Transmembrane  110 - 126 ( 110 - 126)
   INTEGRAL    Likelihood = -0.37   Transmembrane   13 -  29 (  13 -  29)
   PERIPHERAL  Likelihood =  0.47      65
 modified ALOM score:   2.39

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00610(328 - 681 of 981)
SP|Q02009|YTRP_LACLA(1 - 119 of 119) HYPOTHETICAL 13.3 KDA PROTEIN IN TRPE 5'REGION.
GP|551879|gb|AAA25222.1||M87483 ORF 1 {Lactococcus lactis} PIR|S35123|S35123 hypothetical
protein (trpE 5' region) - Lactococcus lactis subsp. lactis
%Match = 19.9
%Identity = 58.8  %Similarity = 77.3
Matches = 70  Mismatches = 26  Conservative Sub.s = 22

114       144       174       204       234       264       294       324
SPKFFQTIQPFLPMTYSVSGLRETISLTGDVNHQWRMLVIFLVSSMILALLIYRKQED**KVSSDRLTV*YGMSKYLGGE

354       384       414       444       474       504       534       561
DMSTLTIIIATLTALEHFYIMYLETLATQSNMTGKIFSMSKEELSYLPVIKLFKNQGVYNGLIGLFLLYGLYISQNQ-EI
 |: ||||:: | ||| ||||||||:|| |  | ::|:| ||||    |  ||||||:||||||| |:| :: |  | ||
 MTILTIILSLLVALEFFYIMYLETFATSSKTTSRVFNMGKEELERSSVQTLFKNQGIYNGLIGLGLIYAIFFSSAQLEI
         10        20        30        40        50        60        70

591       621       651       681       711       741       771       801
VAVFLINVLLVAIYGALTVDKKILLKQGGLPILALLTFLF*YYLAVRFS*TAFSNHFFLIIQVV*VICL*K*YNITTNSK
| ::|| ::|||:||:|| :|||:| |||| ||||:: :|
VRLLLIYIILVALYGSLTSNKKIILTQGGLAILALISSFF
         90       100       110
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1046

A DNA sequence (GBSx1118) was identified in *S.agalactiae* <SEQ ID 3233> which encodes the amino acid sequence <SEQ ID 3234>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3140(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10285> which encodes amino acid sequence <SEQ ID 10286> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12447 GB:Z99107 similar to arylesterase [Bacillus subtilis]
 Identities = 37/91 (40%), Positives = 56/91 (60%)

Query: 13   KDGSDIYYRVVGQGQPIVFLHGNSLSSRYFDKQIAYFSKYYQVIVMDSRGHGKSHAKLNT 72
            +D + +YY   G G PI+F+HG  +S ++F KQ +  S  YQ I +D RGHG+S   L+
Sbjct: 7    EDQTRLYYETHGSGTPILFIHGVLMSGQFFHKQFSVLSANYQCIRLDLRGHGESDKVLHG 66

Query: 73   ISFRQIAVDLKDILVHLEIDKVILVGHSDGA 103
             +  Q A D+++ L  +E+D V+L G S GA
Sbjct: 67   HTISQYARDIREFLNAMELDHVVLAGWSMGA 97
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1047

A DNA sequence (GBSx1119) was identified in *S.agalactiae* <SEQ ID 3235> which encodes the amino acid sequence <SEQ ID 3236>. This protein is predicted to be an integral membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.90    Transmembrane   14 -  30 (   9 -  41)
   INTEGRAL    Likelihood = -9.71    Transmembrane  451 - 467 ( 447 - 472)
   INTEGRAL    Likelihood = -9.18    Transmembrane  234 - 250 ( 229 - 257)
   INTEGRAL    Likelihood = -8.07    Transmembrane   56 -  72 (  46 -  77)
   INTEGRAL    Likelihood = -8.01    Transmembrane  490 - 506 ( 484 - 512)
   INTEGRAL    Likelihood = -5.84    Transmembrane  414 - 430 ( 412 - 436)
   INTEGRAL    Likelihood = -4.99    Transmembrane  136 - 152 ( 135 - 159)
   INTEGRAL    Likelihood = -4.14    Transmembrane  213 - 229 ( 211 - 232)
   INTEGRAL    Likelihood = -4.14    Transmembrane  365 - 381 ( 364 - 382)
   INTEGRAL    Likelihood = -2.66    Transmembrane  393 - 409 ( 391 - 412)
   INTEGRAL    Likelihood = -1.06    Transmembrane  168 - 184 ( 167 - 184)
   INTEGRAL    Likelihood = -0.64    Transmembrane  275 - 291 ( 275 - 291)
   INTEGRAL    Likelihood = -0.32    Transmembrane  328 - 344 ( 328 - 345)
   INTEGRAL    Likelihood = -0.27    Transmembrane  821 - 837 ( 821 - 837)

----- Final Results -----
               bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10283> which encodes amino acid sequence <SEQ ID 10284> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA24464 GB:D85082 YfiX [Bacillus subtilis]
 Identities = 190/596 (31%), Positives = 324/596 (53%), Gaps = 31/596 (5%)
```

```
Query: 246 IVSLIPGGLGSFELVLFTGFAAEGLPKETVVAWLLLYRLAYYIIPFFAGIYFFIHYLGSQ 305
           ++SL+PGG GSF+L+   G    G  +E +V  ++LYRLAY  IPF  G++F    L
Sbjct: 1   MISLVPGGFGSFDLLFLLGMEQLGYHQEAIVTSIVLYRLAYSFIPFILGLFFAAGDLTEN 60

Query: 306 INQRYENVPK-----ELVSTVLQTMVSHLMRILG---AFLIFSTAFFENITYIMWLQKLG 357
             +R E  P+     E  + +L    + L+RIL    + ++F        +  + + +L
Sbjct: 61  TMKRLETNPRIAPAIETTNVLLVVQRAVLVRILQGSLSLIVFVAGLIVLASVSLPIDRLT 120

Query: 358 LDP-LQEQMLWQFPGLLLGVCFILLARTID--QKVKNAFPIAIIWITLTLFYLNLGHISW 414
           + P +    L  F GL L    ILL   I+  ++ K ++ +AI  +     +  L  ++
Sbjct: 121 VIPHIPRPALLLFNGLSLSSALILLILPIELYKRTKRSYTMAITALVGGFVFSFLKGLNI 180

Query: 415 RLSFWFILLLLGLLVIKPTLYKKQFIYSWEERIKDGIIIVSLMGVLFY----IAGLLFPI 470
              F   ++++ L+++K    ++Q  Y+  + I      V+L  V  +    IAG ++
Sbjct: 181 SAIFVLPMIIVLLVLLKKQFVREQASYTLGQLI----FAVALFTVALFNYNLIAGFIWDR 236

Query: 471 RAHITGGSIERLHYIIAWEPIALATL----ILTLVYLCLVKILQGKSCQIGDVFNVDRYK 526
              +    +   +++ +   I  AT+    I+ L +L    +   ++  IG+  + +R
Sbjct: 237 MKKV----LRHEYFVHSTSHITHATIMAIIIVPLFFLIFTVVYHKRTKPIGEKADPERLA 292

Query: 527 KLLQAYGGSSDSGLAFLNDKRLYWYQKNGEDCVAFQFVIVNNKCLIMGEPAGDDTYIREA 586
             L   GG++ S L FL DKR Y +  +G   + F  +    + +++G+P+G
Sbjct: 293 AFLNEKGGNALSHLGFLGDKRFY-FSSDGNALLLFGKIA--RRLVVLGDPSGQRESFPLV 349

Query: 587 IESFIDDADKLDYDLVFYSIGQKLTLLLHEYGFDFMKVGEDALVNLETFTLKGNKYKPFR 646
           +E F+++A +  + ++FY I ++   L H++G++F K+GE+A V+L TFTL G K    R
Sbjct: 350 LEEFLNEAHQKGFSVLFYQIEREDMALYHDFGYNFFKLGEEAYVDLNTFTLTGKKKAGLR 409

Query: 647 NALNRVEKDGFYFEVVQSPHSQELLNSLEEISNTWLEGRPEKGFSLGYFNKDYFQQAPIA 706
              NR E++ + F V   P S   L  L++IS+ WL  + EKGFSLG+F+  Y Q+APIA
Sbjct: 410 AINNRFEREEYTFHVDHPPFSDAFLEELKQISDEWLGSKKEKGFSLGFFDPSYLQKAPIA 469

Query: 707 LVKNAEHEVVAFANIMPNYEKSIISIDLMRHDKQKIPNGVMDFLFLSLFSYYQEKGYHYF 766
            +KNAE E+VAFAN+MP Y++  IS+DLMR+ +   PNG+MD LF+ +F + +E+G   F
Sbjct: 470 YMKNAEGEIVAFANVMPMYQEGEISVDLMRY-RGDAPNGIMDALFIRMFLWAKEEGCTSF 528

Query: 767 DLGMAPLSGVGRVETSFAKERMAYLVYHFGSHFYSFNGLHKYKKKFTPLWSERYIS 822
           ++GMAPL+ VG   TSF  ER A ++++   + YSF+GL  +K+K+ P W  +Y++
Sbjct: 529 NMGMAPLANVGTAFTSFWSERFAAVIFNNVRYMYSFSGLRAFKEKYKPEWRGKYLA 584
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8721> and protein <SEQ ID 8722> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      9.22
GvH: Signal Score (-7.5): -7.66
     Possible site: 58
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 14 value: -12.90 threshold:  0.0
      INTEGRAL    Likelihood =-12.90   Transmembrane   14 -  30 (   9 -  41)
      INTEGRAL    Likelihood = -9.71   Transmembrane  451 - 467 ( 447 - 472)
      INTEGRAL    Likelihood = -9.18   Transmembrane  234 - 250 ( 229 - 257)
      INTEGRAL    Likelihood = -8.07   Transmembrane   56 -  72 (  46 -  77)
      INTEGRAL    Likelihood = -8.01   Transmembrane  490 - 506 ( 484 - 512)
      INTEGRAL    Likelihood = -5.84   Transmembrane  414 - 430 ( 412 - 436)
      INTEGRAL    Likelihood = -4.99   Transmembrane  136 - 152 ( 135 - 159)
      INTEGRAL    Likelihood = -4.14   Transmembrane  213 - 229 ( 211 - 232)
      INTEGRAL    Likelihood = -4.14   Transmembrane  365 - 381 ( 364 - 382)
      INTEGRAL    Likelihood = -2.66   Transmembrane  393 - 409 ( 391 - 412)
      INTEGRAL    Likelihood = -1.06   Transmembrane  168 - 184 ( 167 - 184)
      INTEGRAL    Likelihood = -0.64   Transmembrane  275 - 291 ( 275 - 291)
      INTEGRAL    Likelihood = -0.32   Transmembrane  328 - 344 ( 328 - 345)
      INTEGRAL    Likelihood = -0.27   Transmembrane  821 - 837 ( 821 - 837)
      PERIPHERAL  Likelihood =  1.06     558
```

```
 modified ALOM score:   3.08

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00608(967 - 2787 of 3141)
OMNI|NT01BS0989(20 - 633 of 652) putative integral membrane protein, putative
%Match = 14.6
%Identity = 33.0  %Similarity = 58.0
Matches = 201  Mismatches = 244  Conservative Sub.s = 153

825       855       885       915       945       975       1005      1035
YYLVLIGASMYFPVIYWISGHKGSHYFGDMPSSTRIKLGVVSFFEWGCAAAAFIIIGYLMGIHLPVYKILPLFCIGCAVG
                                                         : ||||  :  :: :| |    |
                                      LELQLLNGSWPGPVIYFALFAMGIHADIRYVFGVFVIAAIGG
                                              10        20        30        40

1065      1095      1125      1155      1185      1215      1245      1260
IVSLIPGGLGSFELVLFTGFAAEGLPKETVVAWLLLYRLAYYIIPFFAGIYFFIHYLGSQINQRYENVPK-----ELVST
::||:|||:|||:|::: |    |  :| :|  ::||||||  |||  |::|    |     :| |  |:     |  :
MISLVPGGFGSFDLLFLLGMEQLGYHQEAIVTSIVLYRLAYSFIPFILGLFFAAGDLTENTMKRLETNPRIAPAIETTNV
                60        70        80        90       100       110       120

1290      1311      1341      1371      1398      1428      1458      1482
VLQTMVSHLMRIL-GAF--LIFSTAFFENITYIMWLQKLGLDP-LQEQMLWQFPGLLLGVCFILLARTID--QKVKNAFP
:|    : |:||| |::  ::|   :    :  : : :| : | :    |  | || |   :|||   |:  :: | ::
LLVVQRAVLVRILQGSLSLIVFVAGLIVLASVSLPIDRLTVIPHIPRPALLLFNGLSLSSALILLILPIELYKRTKRSYT
               140       150       160       170       180       190       200

1512      1542      1572      1602      1632      1659      1689      1719
IAIIWITLTLFYLNLGHISWRLSFWFILLLLGLLVIKPTLYKKQFIYSWEE-RIKDGIIIVSLMGVLFYIAGLLFPIRAH
:||  :   : :  |  ::  :|  |:| :: :|::   | ||||:       :   |  |:|  | ::   |:
MAITALVGGFVFSFLKGLN--ISAIFVLPMIIVLLV---LLKKQFVREQASYTLGQLIFAVALFTVALFNYNLIAGFIWD
               220         230          240       250       260       270

1749      1779      1797      1827      1857      1887      1917      1947
ITGGSIERLHYIIAWEPIALAT----LILTLVYLCLVKILQGKSCQIGDVFNVDRYKKLLQAYGGSSDSGLAFLNDKRLY
     :   ::: :   |  ||    :|: | :| :  : : ::  ||:  : :|   :|   ||:: | | || |||:|
RMKKVLRHEYFVHSTSHITHATIMAIIIVPLFFLIFTVVYHKRTKPIGEKADPERLAAFLNEKGGNALSHLGFLGDKRFY
          290       300       310       320       330       340       350

1977      2007      2037      2067      2097      2127      2157      2187
WYQKNGEDCVAFQFVIVNNKCLIMGEPAGDDTYIREAIESFIDDADKLDYDLVFYSIGQKLTLLLHEYGFDFMKVGEDAL
 :  :|   : |    :  : :::|:|:|        :| |:::| :  :  ::|| | ::   | |::|::| |:||:|
-FSSDGNALLLF--GKIARRLVVLGDPSGQRESFPLVLEEFLNEAHQKGFSVLFYQIEREDMALYHDFGYNFFKLGEEAY
             370       380       390       400       410       420       430

2217      2247      2277      2307      2337      2367      2397      2427
VNLETFTLKGNKYKPFRNALNRVEKDGFYFEVVQSPHSQELLNSLEEISNTWLEGRPEKGFSLGYFNKDYFQQAPIALVK
|:| |||| | |   :|   || |:: : | | : | |  :|  |::||: ||  : |||||||:|:  |:|:|||| :|
VDLNTFTLTGKKKAGLRAINNRFEREEYTFHVDHPPFSDAFLEELKQISDEWLGSKKEKGFSLGFFDPSYLQKAPIAYMK
             450       460       470       480       490       500       510

2457      2487      2517      2547      2577      2607      2637      2667
NAEHEVVAFANIMPNYEKSIISIDLMRHDKQKIPNGVMDFLFLSLFSYYQEKGYHYFDLGMAPLSGVGRVETSFAKERMA
||| |:|||||:|| |::  ||:||||: :   |||:|| ||: :| : :|:|   |::|||||: ||   |||  || |
NAEGEIVAFANVMPMYQEGEISVDLMRY-RGDAPNGIMDALFIRMFLWAKEEGCTSFNMGMAPLANVGTAFTSFWSERFA
             530       540        550       560       570       580       590

2697      2727      2757      2787      2817      2847      2877      2907
YLVYHFGSHFYSFNGLHKYKKKFTPLWSERYISCSRSSWLICAICALLMEDSKIKIVK*ALFGN*KEHVMRHALFKSFNT
 ::::   : |||:||  :|:|: | |  :|::  ::  |
```

```
AVIFNNVRYMYSFSGLRAFKEKYKPEWRGKYLAYRKNRSLSVTMFLVTRLIGKSKKDSV
              610       620       630       640       650
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1048

A DNA sequence (GBSx1120) was identified in *S.agalactiae* <SEQ ID 3237> which encodes the amino acid sequence <SEQ ID 3238>. This protein is predicted to be choline transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.24   Transmembrane   28 -  44 (  22 -  47)
   INTEGRAL    Likelihood = -8.81   Transmembrane  178 - 194 ( 176 - 204)
   INTEGRAL    Likelihood = -7.22   Transmembrane   81 -  97 (  63 - 105)
   INTEGRAL    Likelihood = -3.50   Transmembrane  209 - 225 ( 206 - 226)
   INTEGRAL    Likelihood = -3.13   Transmembrane   64 -  80 (  63 -  80)
   INTEGRAL    Likelihood = -2.44   Transmembrane  156 - 172 ( 153 - 172)
   INTEGRAL    Likelihood = -0.64   Transmembrane  137 - 153 ( 137 - 153)

----- Final Results -----
              bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD45530 GB:AF162656 choline transporter [Streptococcus pneumoniae]
 Identities = 326/505 (64%), Positives = 409/505 (80%), Gaps = 1/505 (0%)

Query: 1   MTTLITTFQERFGDWTQSLIEHLQLSLLTLILATLIAIPLGIIISHYKKISHVVLQITGI 60
           MT LI TFQ+RF DW  +L +HLQLSLLTL+LA L+AIPL + + +++K++  VLQI GI
Sbjct: 1   MTNLIATFQDRFSDWLTALSQHLQLSLLTLLLAILLAIPLAVFLRYHEKLADWVLQIAGI 60

Query: 61  FQTIPSLALLGLFIPFMGIGTVPAVVALIIYALFPILQNTVTVLMQIDANLIEAATAFGM 120
           FQTIPSLALLGLFIP MGIGT+PA+ AL+IYA+FPILQNT+T L  ID NL EA  AFGM
Sbjct: 61  FQTIPSLALLGLFIPLMGIGTLPALTALVIYAIFPILQNTITGLKGIDPNLQEAGIAFGM 120

Query: 121 TRWERLKKFELALSMPVIISGIRTASVMIIGTATLASLIGAGGLGSFILLGIDRNNPSLI 180
           TRWERLKKFE+ L+MPVI+SGIRTA+V+IIGTATLA+LIGAGGLGSFILLGIDRNN SLI
Sbjct: 121 TRWERLKKFEIPLAMPVIMSGIRTAAVLIIGTATLAALIGAGGLGSFILLGIDRNNASLI 180

Query: 181 LIGAISSAVLAIIFSGLIGLLEKARLRTIAVSGILLLAGLGLSYAPKWMPGTNTATITVA 240
           LIGA+SSAVLAI F+ L+ ++EKA+LRTI     L+   LGLSY+P  +       + +A
Sbjct: 181 LIGALSSAVLAIAFNFLLKVMEKAKLRTIFSGFALVALLLGLSYSPALLVQKEKENLVIA 240

Query: 241 GKLGTEPDILINMYKELIEDQTDIKVKLKPNFGKTTFLYQALKSGDIDLYPEFTGTITSS 300
           GK+G EP+IL NMYK LIE+ T +   +KPNFGKT+FLY+ALK GDID+YPEFTGT+T S
Sbjct: 241 GKIGPEPEILANMYKLLIEENTSMTATVKPNFGKTSFLYEALKKGDIDIYPEFTGTVTES 300

Query: 301 LLKNPPKVSNNPKQVYNLAKNGILKQDKLSLLSPMAYQNTYAVAVKKDYAEANQLKNISD 360
           LL+  PKVS+ P+QVY +A++GI KQD L+ L PM+YQNTYAVAV K  A+   LK ISD
Sbjct: 301 LLQPSPKVSHEPEQVYQVARDGIAKQDHLAYLKPMSYQNTYAVAVPKKIAQEYGLKTISD 360

Query: 361 LKKLD-KLKAGFTLEFKDREDGSIGLQKHYGLNLDISTLEPALRYQAINSKDVNIIDAYS 419
           LKK++ +LKAGFTLEF DREDG+ GLQ  YGLNL+++T+EPALRYQAI S D+ I DAYS
Sbjct: 361 LKKVEGQLKAGFTLEFNDREDGNKGLQSMYGLNLNVATIEPALRYQAIQSGDIQITDAYS 420

Query: 420 TDSELIQYQLQILKDDKHLFPPYQGAPLLRQDTIKKYPQVKKALNKLAGHITEKEMQEMN 479
           TD+EL +Y LQ+L+DDK LFPPYQGAPL+++  +KK+P++++ LN LAG ITE +M ++N
Sbjct: 421 TDAELERYDLQVLEDDKQLFPPYQGAPLMKEALLKKHPELERVLNTLAGKITESQMSQLN 480

Query: 480 YQVAVKHKSAATVAKQYLKAHHIIK 504
```

```
           YQV V+ KSA  VAK++L+   ++K
Sbjct: 481 YQVGVEGKSAKQVAKEFLQEQGLLK 505
```

There is also homology to SEQ ID 636.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1049

A DNA sequence (GBSx1121) was identified in *S.agalactiae* <SEQ ID 3239> which encodes the amino acid sequence <SEQ ID 3240>. This protein is predicted to be choline transporter (opuBA). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2345(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD45529 GB:AF162655 choline transporter [Streptococcus pneumoniae]
 Identities = 139/236 (58%), Positives = 178/236 (74%)

Query: 1   MISFENVSKSYGDHTIIDNISCHIQRGEFFVLVGASGSGKTTILKMINRLIEPSQGAITL 60
           MI ++NV+  Y +  ++ +++  I+ GEF VLVG SGSGKTT+LKMINRL+EP+ G I +
Sbjct: 1   MIEYKNVALRYTEKDVLRDVNLQIEDGEFMVLVGPSGSGKTTMLKMINRLLEPTDGNIYM 60

Query: 61  DGENITSLDLRQLRLETGYVLQQIALFPNLTVGENIELIPEMKGWSKGDQKKAASDLLDK 120
           DG+ I   D R+LRL TGYVLQ IALFPNLTV ENI LIPEMKGWSK +  K   +LL K
Sbjct: 61  DGKRIKDYDERELRLSTGYVLQAIALFPNLTVAENIALIPEMKGWSKEEITKKTEELLAK 120

Query: 121 VGLPAKDYFNRYPHELSGGEQQRIGILRAIVAKPKVLLMDEPFSALDPISRRQLQDITKQ 180
           VGLP  +Y +R P ELSGGEQQR+GI+RA++ +PK+ LMDEPFSALD ISR+QLQ +TK+
Sbjct: 121 VGLPVAEYGHRLPSELSGGEQQRVGIVRAMIGQPKIFLMDEPFSALDAISRKQLQVLTKE 180

Query: 181 LQSELGITLVFVTHDMKEAMRLADRICVIKEGKIVQLDRPEIIQNNPSDQFVRTLF 236
           L  E G+T +FVTHD  EA++LADRI V+++G+I Q+  PE I   P+  FV  LF
Sbjct: 181 LHKEFGMTTIFVTHDTDEALKLADRIAVLQDGEIRQVANPETILKAPATDFVADLF 236
```

There is also homology to SEQ ID 644.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1050

A DNA sequence (GBSx1122) was identified in *S.agalactiae* <SEQ ID 3241> which encodes the amino acid sequence <SEQ ID 3242>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.52   Transmembrane   49 -  65 (  46 -  66)

----- Final Results -----
               bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06434 GB:AP001516 two-component response regulator [Bacillus halodurans]
 Identities = 101/305 (33%), Positives = 152/305 (49%), Gaps = 31/305 (10%)

Query: 1   MKFYIIDDDPTITMILQDIIE-EDFNNTVVRVNNVSSKAYNELLIADVDIVLIDLLMPIL 59
           M F+I DDD T+  IL  IIE E     V    + S      L I  VDI+LIDLLMP
Sbjct: 1   MNFFITDDDVTVRSILAQIIEDEQLGQVVGEAEDGSELDGKRLNIKQVDILLIDLLMPNC 60

Query: 60  DGVTLVQKIYKQRSDLKFIMISQVKDNDLRQEAYKAGIEFFINKPINIIEVKSVVKRVTD 119
           DG+  +QKI K     K IMISQ++  +L  EAY  GIE +I KPIN IEV SV+++V +
Sbjct: 61  DGLEAIQKI-KPEFKGKIIMISQIESKELISEAYLLGIEHYIMKPINKIEVLSVIRKVIN 119

Query: 120 TIEMQKKLNTIQNLLENTPSYQKPITTSNLT----KIRS----ILSYLGITSETAYTDIL 171
              +++ L  IQ  L N      P   ++       I+S    +LS LGI  E+   D++
Sbjct: 120 HTRLEQSLYDIQKSLSNVLQGSIPTQVNDQVFHDDSIKSYGQYLLSELGIAGESGSKDLM 179

Query: 172 NICELLLKQELNF-------AQFDFQKELSIDE-----------HQQKIILQRIRRAVKK 213
           NI   L   E  +       A  D  ++L+ ++            + K   QR+RRAV +
Sbjct: 180 NILMFLYTYEKEYSFEKGFPALKDIFEQLASEKLGDAADERDVRREVKAAKQRVRRAVYQ 239

Query: 214 AMINMAHLYIDDFENELTLQYANALFGFQNIHNEAQLIQGK---SMYGGKISLKHFFDEL 270
           ++ ++A L + DF N    +YA+  F F  + ++   ++ +   S    +I++K F   L
Sbjct: 240 SLEHVASLGLIDFSNPKFEEYASHFFDFSVVRSKMTELKNETSSSYTSARINVKKFTQAL 299

Query: 271 ILQSK 275
             ++K
Sbjct: 300 YYEAK 304
```

There is homology to SEQ ID 460.

A related GBS gene <SEQ ID 8723> and protein <SEQ ID 8724> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -7.05
GvH: Signal Score (-7.5): -6.58
     Possible site: 61
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -5.52 threshold:  0.0
   INTEGRAL    Likelihood = -5.52   Transmembrane   49 -  65 (  46 -  66)
   PERIPHERAL  Likelihood =  7.37     155
 modified ALOM score:   1.60

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00604(307 - 1125 of 1431)
EGAD|137180|146289(3  -  304  of  310)  hypothetical  protein  {Bacillus  cereus}
GP|1769946|emb|CAA67094.1||X98455 orf1 {Bacillus cereus}
%Match = 12.7
%Identity = 34.1  %Similarity = 53.0
Matches = 95  Mismatches = 123  Conservative Sub.s = 53

168       198       228       258       288       318       348       375
*C*W*YLSRNRAIPRAYFNGRAISRNDNCLS*SAKWNNIYTVIP*KSI*VRR*YVKFYIIDDDPTITMILQDIIEE-DFN
                                                        :||:|||     :|  |||: |:
                                                      MFYYIVDDDEVFRSMLSQIIEDGDLG
                                                              10        20
```

```
405       435       465       495       525       555       585       615
NTVVRVNNVSSKAYNELLIADVDIVLIDLLMPILDGVTLVQKIYKQRSDLKFIMISQVKDNDLRQEAYKAGIEFFINKPI
  :    : :     :|    |||::||||||: ||:  |: |       | ||||||:    |  |||  |:|::| ||:
EVIGESEDGAFVEAEQLNYKKVDILFIDLLMPMRDGIETVRHI-ASSFTGKIIMISQVESKQLIGEAYTLGVEYYITKPL
             40        50        60        70        80        90        100

645       675       705                 753       771       801       831
NIIEVKSVVKRVTDTIEMQKKLNTIQNLLENTPSYQKP----ITTSNLTKI----RSILSYLGITSETAYTDILNICELL
| ||| |||::| : | ::: :  ||  | |   ::||     |     ||    | :|: |||  |    |:|:: | |
NKIEVVSVVRKVIERIRLERSIYDIQKSLNNVFQWEKPQMRSETVQEEKKISDSGRFLLAELGIAGENGSKDLLSMLEYL
             120       130       140       150       160       170       180

861                           894       924       954       984      1014
LKQELNFAQFDFQKELSID-------------------EHQQKIILQRIRRAVKKAMINMAHLYIDDFENELTLQYANAL
  ||     |:|      |                   | ::|   ||:|||: ::: ::| | : || |     ||  :
YGQE-KAQTFEFGFPALKDIFHQITLKKLGEIASDADIEKEKKASEQRVRRAIYQSLNHLASLGLTDFSNPKFESYAPKF
              200       210       220       230       240       250       260

1071      1095      1125      1155      1185      1215      1245
FGFQNIHNE-AQLIQGKSMYGGKISL--KHFFDELILQSKTF*DLFKHGLIYYNHPKTFLFINLQQTPCLPQGVCFCF*F
| |  :      :: :       | | :  | |   | :::|
FDFTVVRKRMTEMTKDGVATSGHIRINTKKFIQVLYFEAKRLMEIE
              280       290       300       310
```

SEQ ID 8724 (GBS356) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 3; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 8; MW 59kDa).

GBS356-GST was purified as shown in Figure 216, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1051

A DNA sequence (GBSx1123) was identified in *S.agalactiae* <SEQ ID 3243> which encodes the amino acid sequence <SEQ ID 3244>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.48    Transmembrane  149 - 165 ( 147 - 172)
   INTEGRAL    Likelihood = -5.20    Transmembrane   37 -  53 (  29 -  55)
   INTEGRAL    Likelihood = -2.50    Transmembrane  126 - 142 ( 126 - 142)
   INTEGRAL    Likelihood = -2.13    Transmembrane   62 -  78 (  60 -  78)
   INTEGRAL    Likelihood = -0.64    Transmembrane  314 - 330 ( 314 - 330)
   INTEGRAL    Likelihood = -0.11    Transmembrane   89 - 105 (  89 - 105)

----- Final Results -----
               bacterial membrane --- Certainty=0.3590(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06435 GB:AP001516 two-component sensor histidine kinase
          [Bacillus halodurans]
 Identities = 118/427 (27%), Positives = 199/427 (45%), Gaps = 25/427 (5%)

Query: 10   LERRQRIIISAIAIA-LAAQINISILADGFIMTLSLFILPVFLYFNDDINPILLCLGITF 68
            L +   II+S +  A +A +IN   + + F ++L   I  +FL F  +   I+
Sbjct: 7    LSKDYMIILSMLLFAPIAGEINFYPVNETFRVSLGPPIFFLFLLFLRNTAAIVPGFFTAI 66

Query: 69   ASPIFRGIILSIAGEAEIHQIIEFVLTDMAFYICYGITFYTIYWHRSYRNKGTFFFSIII 128
            A  +FR  + ++  +       E       FY  Y + F      R +      F   II
Sbjct: 67   AVVVFRVFLDTLHADFYWVDSFEIHYPTFFFYFTYSLLFSLAKVQRFHEQPLIIFLFGII 126
```

```
Query: 129 CDYFANLVEISFLIKFNNYTITIFA-TLFAIALLRAFISCAVAYTYSYLSLLLQKD---D 184
            +  A+  E  F+ ++  + + +   ++F I L+    S  V   +S + L   +    +
Sbjct: 127 IEILADTAE--FIAQYFAFGVMVTKDSIFQILLIAFSHSFIVLGVFSMMKLYETRSRELE 184

Query: 185 HERRYYYFMWSTSAVKSEVYFMQKNIIEIENIMKNAYLLDKELSKY---HLPKEYQHLS- 240
             +R  + +   S +  E   ++K +   E+I    + L +E+ +    H+ +   HL
Sbjct: 185 IRKRNEHMLLLISNLYEESVHLKKTLQNSEDITSKVFGLYREMKRLQSEHMDQVNPHLEK 244

Query: 241 -----LDISRDVHEVKKDYQNIIKGLGTYFSVKNESTMALKDIFQIVLSYTRS---IIQF 292
                L+IS +VHE+KKD Q I  GL    S  NES +   +I QI+    R+     Q
Sbjct: 245 ISKRLLEISGEVHEIKKDNQRIFAGLSKLIS--NESYVDYIEIGQIIKMIVRTNEKYAQL 302

Query: 293 RHQDIIILENNKCNLIISNYYYLLTIISNIVLNAVEAIDKQKKGTISVHTEELEDFIKIE 352
             ++I    + +      + Y  L+II+N+V NAVEAID   KG +++  + L   ++
Sbjct: 303 LGKEIDFHYSIQGEHPPYHIYTHLSIINNLVANAVEAIDG--KGMLTIRVKALGQTVEFR 360

Query: 353 ISDNGPGIPDKMKHMIFKPGFSTKFDANGDIYRGIGLSHVRILMEEQYQGTITVCPNQ-P 411
           I D+GPGIPDK + +IFKPGF++KFD  G    GIGL++V   M ++  GT+     Q
Sbjct: 361 IEDDGPGIPDKHRALIFKPGFTSKFDHTGKPSTGIGLTYVHD-MVDKLGGTVVYERGQGG 419

Query: 412 NGTTFTL 418
            G+ FT+
Sbjct: 420 KGSVFTI 426
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1052

A DNA sequence (GBSx1124) was identified in *S.agalactiae* <SEQ ID 3245> which encodes the amino acid sequence <SEQ ID 3246>. This protein is predicted to be ornithine carbamoyltransferase Otc6850 (argF). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane  171 - 187 ( 171 - 187)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB75986 GB:AJ272085 ornithine carbamoyltransferase
           [Staphylococcus aureus]
 Identities = 264/332 (79%), Positives = 292/332 (87%)

Query: 1   MKNLRNRSFLTLLDFSTAEVEFLLKLSEDLKRAKYAGIEQQKLVGKNIALIFEKDSTRTR 60
           MKNLRNRSFLTLLDFS  EVEFLL LSEDLKRAKY G E+  L  KNIAL+FEKDSTRTR
Sbjct: 1   MKNLRNRSFLTLLDFSRQEVEFLLTLSEDLKRAKYIGTEKPMLKNKNIALLFEKDSTRTR 60

Query: 61  CAFEVAAHDQGAHVTYLGPTGSQMGKKETSKDTARVLGGMYDGIEYRGFSQETVETLAEF 120
           CAFEVAAHDQGA+VTYLGPTGSQMGKKET+KDTARVLGGMYDGIEYRGFSQ TVETLAE+
Sbjct: 61  CAFEVAAHDQGANVTYLGPTGSQMGKKETTKDTARVLGGMYDGIEYRGFSQRTVETLAEY 120

Query: 121 SGVPVWNGLTDADHPTQVLADFLTAKECLHKPYKDIRFTYVGDGRNNVANALMIGASIVG 180
           SGVPVWNGLTD DHPTQVLADFLTAKE L K Y DI FTYVGDGRNNVANALM GA+I+G
Sbjct: 121 SGVPVWNGLTDEDHPTQVLADFLTAKEVLKKDYADINFTYVGDGRNNVANALMQGAAIMG 180

Query: 181 MTYHLVCPKELEPDPELLSKCQEIAKTTGASIEITADIAEGVRDSDVLYTDVWVSMGEPD 240
           M +HLVCPKEL P  ELL++C+ IA   G +I IT DI +GV+ SDV+YTDVWVSMGEPD
Sbjct: 181 MNFHLVCPKELNPTDELLNRCKNIAAENGGNILITDDIDQGVKGSDVIYTDVWVSMGEPD 240
```

```
Query: 241 EVWKERIALLEPYRITQEMLNMTENPNVIFEHCLPSFHNIDTKVGYDIYEKYGLKEMEVS 300
           EVWKER+ LL+PY++ +EM++ T NPNVIFEHCLPSFHN DTK+G  I+EKYG++EMEV+
Sbjct: 241 EVWKERLELLKPYQVNKEMMDKTGNPNVIFEHCLPSFHNADTKIGQQIFEKYGIREMEVT 300

Query: 301 DEVFEGPHSVVFQEAENRMHTIKAVMVATLGD 332
           DEVFE   SVVFQEAENRMHTIKAVMVATLG+
Sbjct: 301 DEVFESKASVVFQEAENRMHTIKAVMVATLGE 332
```

There is also homology to SEQ ID 3118.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1053

A DNA sequence (GBSx1126) was identified in *S.agalactiae* <SEQ ID 3247> which encodes the amino acid sequence <SEQ ID 3248>. This protein is predicted to be carbamate kinase (b2874). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.48   Transmembrane  214 - 230 ( 214 - 230)

----- Final Results -----
               bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA66367 GB:X97768 carbamate kinase [Clostridium perfringens]
 Identities = 162/313 (51%), Positives = 207/313 (65%), Gaps = 7/313 (2%)

Query: 3   KIVVALGGNAL-----GNSPEEQLRLVKHTAKSLVALIKKGHEIVVSHGNGPQVGAINLG 57
           KIV+ALG NAL       S E QL   + TA S+  LI+ GHE+ + HGNGPQVG I
Sbjct: 2   KIVLALGENALQKDSKDKSAEGQLETCRQTAISVADLIEDGHEVSIVHGNGPQVGQILAS 61

Query: 58  MNFAAESGQGTN-FPFPECGAMSQGYIGYHLQQSLLNELRQEGINKEVATIITQIEVDES 116
           +  A +   G   FPF   GA S+GYIGYHLQ ++  EL + GI K V TI TQ+ VD++
Sbjct: 62  IELAHQVDNGNPLFPFDVVGAFSEGYIGYHLQNTIREELLKRGIEKSVDTITTQVIVDKN 121

Query: 117 DQAFSAPTKPIGTFYDKETSEKIAIEKGYTFVEDAGRGYRRVVASPEPKKIIEINSIKTL 176
           D  F+ PTKPIG+FY KE +EK+  +KGYT  EDAGRGYRRVVASP+P  I+E  +IKT+
Sbjct: 122 DPGFTNPTKPIGSFYTKEEAEKLEKDKGYTMKEDAGRGYRRVVASPKPVDIVEKEAIKTM 181

Query: 177 IENDTLVIAGGGGGIPVINKGG-YEGIAAVIDKDKSSALLAGELAADQLIILTAVDYVYT 235
           +++  +VIA GGGGIPV+  G   EG+ AVIDKD ++  LA  L AD L+ILTAVD V
Sbjct: 182 VDSGFIVIACGGGGIPVVEDGDRLEGVPAVIDKDFAAEKLAEILDADALLILTAVDRVCV 241

Query: 236 QFGKENQKALTEVNENQMIDYVNQGEFAKGSMLPKVIACMSFLDHNPKGTALITSLNGLE 295
            F K +QKAL E+N  ++  Y+ +G+FA GSMLPKV AC  F+    K  A+I SL   +
Sbjct: 242 NFNKPDQKALKEINLEEVDKYIEEGQFAPGSMLPKVEACKKFVLSGDKKVAIIASLTNAK 301

Query: 296 DALDGKLGTRITK 308
            AL G+ GT+I K
Sbjct: 302 AALRGESGTKIVK 314
```

There is also homology to SEQ ID 3110.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1054

A DNA sequence (GBSx1127) was identified in *S.agalactiae* <SEQ ID 3249> which encodes the amino acid sequence <SEQ ID 3250>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3558(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1055

A DNA sequence (GBSx1128) was identified in *S.agalactiae* <SEQ ID 3251> which encodes the amino acid sequence <SEQ ID 3252>. This protein is predicted to be a transmembrane protein (b2298). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.11   Transmembrane  413 - 429 ( 405 - 440)
   INTEGRAL    Likelihood = -9.61   Transmembrane  498 - 514 ( 489 - 516)
   INTEGRAL    Likelihood = -9.45   Transmembrane  165 - 181 ( 161 - 185)
   INTEGRAL    Likelihood = -8.07   Transmembrane  127 - 143 ( 122 - 146)
   INTEGRAL    Likelihood = -7.22   Transmembrane  308 - 324 ( 306 - 326)
   INTEGRAL    Likelihood = -5.57   Transmembrane  334 - 350 ( 330 - 357)
   INTEGRAL    Likelihood = -4.51   Transmembrane  194 - 210 ( 193 - 217)
   INTEGRAL    Likelihood = -3.82   Transmembrane  372 - 388 ( 371 - 390)
   INTEGRAL    Likelihood = -1.22   Transmembrane  250 - 266 ( 250 - 268)
   INTEGRAL    Likelihood = -0.80   Transmembrane  468 - 484 ( 468 - 484)
   INTEGRAL    Likelihood = -0.32   Transmembrane  436 - 452 ( 436 - 452)

----- Final Results -----
               bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22251 GB:U32741 conserved hypothetical transmembrane protein
          [Haemophilus influenzae Rd]
 Identities = 303/506 (59%), Positives = 389/506 (75%), Gaps = 6/506 (1%)

Query: 10  NKRSKGFRMPGAFTILFILTIFSVLATWWIPAGSYSKLQFDTASSKLVVTDPNGKTVHVP 69
           +K+ K F  P AFTILF + I +V  TW IP+GSYSKL +++  +  VV          P
Sbjct: 4   SKKKKTFNFPSAFTILFAILILAVGLTWVIPSGSYSKLTYNSTDNVFVVKAYGVDDKTYP 63

Query: 70  ATQTQLDKMNVKIKIKEFTSGAISKPVSVPNTYKRLKQNPAGIGSVTTSMVNGTIEAVDI 129
           AT   LD +N+KIK+  FT G I KP+++P TY+R++Q+  GI  +T SMV GTIEAVD+
Sbjct: 64  ATTDTLDNLNIKIKLSNFTEGVIKKPIAIPGTYQRVEQHHKGIEDITKSMVEGTIEAVDV 123

Query: 130 MVFIMVLGGMIGVVRKSGAFESGLLALTKKTKGREFLLIFLVSLLMVLGGTLCGIEEEAV 189
           MVFI VLGGMIGV+ ++G+F +GL+AL KKTKG EF ++F VS+LMVLGGT CGIEEEAV
Sbjct: 124 MVFIFVLGGMIGVINRTGSFNAGLMALVKKTKGNEFFIVFCVSVLMVLGGTTCGIEEEAV 183

Query: 190 AFYPILVPIFLAMGYDSIICVGAIFLASSVGTSFSTINPFSSVIASNAAGISFTEGLSWR 249
```

```
           AFYPILVP+FLA+GYD+I+CVGAIFLA+S+GT+FSTINPFS VIASNAAGI FTEG+ +R
Sbjct: 184 AFYPILVPVFLALGYDAIVCVGAIFLAASMGTAFSTINPFSVVIASNAAGIQFTEGIGFR 243

Query: 250 TAGCIAGAIFVVVYLHWYAKKIKANPEFSYSYEDRVEFNAKWGMTTN-HTPSLFTIRQKI 308
             G + GA  V+ YL+WY KKIKA+P FSY+Y+DR EF  ++    + +T   F+ R+K+
Sbjct: 244 ALGLVLGATCVIAYLYWYCKKIKADPSFSYTYDDREEFRQRYMKNFDPNTTIPFSARRKL 303

Query: 309 ILSLFVISFPLMVWGVMSQGWWFPTMASSFLAITIIIMFLTATGANGIGERDVVDEFVNG 368
           IL+LF ISFP+M+WGVM  GWWFP MA+SFLAITIIIMF+     +G+ E+D+++ F  G
Sbjct: 304 ILTLFCISFPIMIWGVMVGGWWFPQMAASFLAITIIIMFI-----SGLSEKDIMESFTEG 358

Query: 369 ASSLVGVSLIIGLARGINIILSQGYISDTMLYTASKLASHVSGSVFIIVMMFIYFVLGFV 428
           AS LVGVSLIIGLARG+N++L QG ISDT+L   S + S + GSVFI+  + ++  LG +
Sbjct: 359 ASELVGVSLIIGLARGVNLVLEQGMISDTILDYMSNVVSGMPGSVFILGQLVVFIFLGLI 418

Query: 429 VPSSSGLAVLSMPILAPLADTVGIPRSVVVMAYQFGQYAMLFLAPTGLVMATLQMLDMKY 488
           VPSSSGLAVLSMPI+APLAD+VGIPR +VV AY +GQYAMLFLAPTGLV+ TLQML + +
Sbjct: 419 VPSSSGLAVLSMPIMAPLADSVGIPRDIVVSAYNWGQYAMLFLAPTGLVLVTLQMLQIPF 478

Query: 489 SHWLKFVWPVVLFLLIFGGGLLVLQV 514
             W+KFV P++  LL+ G  LLV+QV
Sbjct: 479 DRWVKFVMPMIGCLLLIGSILLVVQV 504
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3253> which encodes the amino acid sequence <SEQ ID 3254>. Analysis of this protein sequence reveals the following:

```
         Possible site: 36
>>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood =-13.21   Transmembrane  479 - 495 ( 472 - 496)
     INTEGRAL    Likelihood =-10.24   Transmembrane  261 - 277 ( 258 - 280)
     INTEGRAL    Likelihood = -9.24   Transmembrane  153 - 169 ( 142 - 180)
     INTEGRAL    Likelihood = -7.17   Transmembrane  393 - 409 ( 391 - 411)
     INTEGRAL    Likelihood = -6.00   Transmembrane   81 -  97 (  78 -  99)
     INTEGRAL    Likelihood = -5.95   Transmembrane  318 - 334 ( 314 - 338)
     INTEGRAL    Likelihood = -3.77   Transmembrane  352 - 368 ( 352 - 369)
     INTEGRAL    Likelihood = -2.66   Transmembrane  120 - 136 ( 119 - 138)
     INTEGRAL    Likelihood = -0.32   Transmembrane  204 - 220 ( 204 - 220)

----- Final Results -----
                 bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB94000 GB:AF008219 unknown [Borrelia afzelii]
 Identities = 174/496 (35%), Positives = 306/496 (61%), Gaps = 37/496 (7%)

Query: 10  RIPSSYTVLFIIIAIMAVLTWFIPAGAYETAK---GGG-----VISGTYKTVASNPQGFF 61
           ++PSS+T++F +I  + +LT+ IPAG ++      G G     +++GTY+T+   P+GF
Sbjct: 3   KMPSSFTIIFSLIVFVTILTYVIPAGKFDKEFRQIGDGPKREIIVAGTYQTIDRGPRGFL 62

Query: 62  DILMAPVRGMLGVEGTDGAIQVSFFILMVGGFLGVVNKTGALDTGIASVVRKNKGREKML 121
             +M  +  M   +G + A +V  F+L+VGG  G++ KTGA+D GI S+++K   ++K+L
Sbjct: 63  HPIMTILTAMS--KGMEHAAEVIIFVLIVGGAYGIIMKTGAIDAGIYSLIKKLGHKDKLL 120

Query: 122 IAILIPLFALGGTTYGMGEETMAFYPLLIPVMIAVGFDSIVAVAIILIGSQIGCLASTIN 181
           I +L+ +F++GGT  GM EET+ FY ++IP+++A+G+D++V VAII +G+ +G +AST+N
Sbjct: 121 IPLLMFIFSIGGTVTGMSEETLPFYFVMIPLIVALGYDNVVGVAIIALGAGVGTMASTVN 180

Query: 182 PFATGVAADAAGVSIADGMIWRVIQWVILVGMSIWFVYNYASKIEEDPSKSLVADKEEEH 241
           PFATG+A+  A +S+ DG  +R++ + I + ++I +V  YAS+I++DPSKSLV  K+ EH
Sbjct: 181 PFATGIASAIASISLQDGFSFRIVLYFISILVAIIYVCVYASRIKKDPSKSLVYSKKNEH 240

Query: 242 KELF-QLQNSGEDLNKRQRNVLTIFTLTFVIMILSLIPWEDFGIKFFTNINTWLTTMPIL 300
            + F + + S ED      NV      TF   ++ L+    FG      I   + ++  L
Sbjct: 241 YQYFVKNEISKED------NVQNTLEFTFARKLVLLL----FGFM----ILFLVFSIVQL 286
```

```
Query: 301 GGVIGKTMGAFGTWYFPEITMLFIMMGVLVAIVYRMSEEDFFSSFLTGAGEFLGVAMICA 360
           G            W+  E+TML++ + ++ A + R+ E + + +F+ G+   +  A+I
Sbjct: 287 G------------WWMQEMTMLYLGVAIISAFICRLGESEMWDAFVKGSESLITAALIIG 334

Query: 361 IARGIQVIMNGGMITATILHLGETSLSGLSSQVFVILAYIFYLPMSFLIPSTSGLAGATM 420
           +ARG+ ++ + G+ITAT+L+     L  L    F+IL  I  + + F++PS+SG A  TM
Sbjct: 335 LARGVMIVCDDGLITATMLNAATNFLYNLPRPFFIILNEIIQIFIGFIVPSSSGHASLTM 394

Query: 421 GIMAPLGQFSNVPAHLVITAFQSASGILNMISPTSAIVMGALALGRVDLGTWWKFIGKFI 480
            IMAPL  F ++    V+ A Q++SG++N+I+PTS ++M  L + ++  GTW+KF+
Sbjct: 395 PIMAPLADFLSIGRSSVVIAMQTSSGLINLITPTSGVIMAVLGISKLSYGTWFKFVLPLF 454

Query: 481 VMVMLVSVLLLVVATF 496
           ++   +S+L+++   +
Sbjct: 455 IIEFFISILVIIANVY 470
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 158/542 (29%), Positives = 274/542 (50%), Gaps = 92/542 (16%)

Query: 11  KRSKGFRMPGAFTILFILTIFSVLATWWIPAGSYSKLQFDTASSKLVVTDPNGKTVHVPA 70
           ++ +GFR+P ++T+LFI+     + TW+IPAG+Y     +TA
Sbjct: 4   EKKRGFRIPSSYTVLFIIIAIMAVLTWFIPAGAY-----ETAKG---------------- 42

Query: 71  TQTQLDKMNVKIKIKEFTSGAISKPVSVPNTYKRLKQNPAGIGSVTTSMVNG------TI 124
                               G IS       TYK +  NP G   +  + V G      T
Sbjct: 43  ------------------GGVIS------GTYKTVASNPQGFFDILMAPVRGMLGVEGTD 78

Query: 125 EAVDIMVFIMVLGGMIGVVRKSGAFESGLLALTKKTKGREFLLIFLVSLLMVLGGTLCGI 184
            A+ +  FI+++GG +GVV K+GA ++G+ ++ +K KGRE +LI ++  L  LGGT  G+
Sbjct: 79  GAIQVSFFILMVGGFLGVVNKTGALDTGIASVVRKNKGREKMLIAILIPLFALGGTTYGM 138

Query: 185 EEEAVAFYPILVPIFLAMGYDSIICVGAIFLASSVGTSFSTINPFSSVIASNAAGISFTE 244
            EE +AFYP+L+P+ +A+G+DSI+ V  I + S +G   STINPF++ +A++AAG+S  +
Sbjct: 139 GEETMAFYPLLIPVMIAVGFDSIVAVAIILIGSQIGCLASTINPFATGVAADAAGVSIAD 198

Query: 245 GLSWRTAGCIAGAIFVVVYLHWYAKKIKANPEFSYSYEDRVEFNAKWGMTTNHTPSLFTI 304
           G+ WR    +      + +++ YA KI+ +P  S    D+ E + +     N    L
Sbjct: 199 GMIWRVIQWVILVGMSIWFVYNYASKIEEDPSKSL-VADKEEEHKELFQLQNSGEDL-NK 256

Query: 305 RQKIILSLFVISFPLMV-----W----------------------GVMSQ------GWWF 331
           RQ+ +L++F ++F +M+     W                      GV+ +       W+F
Sbjct: 257 RQRNVLTIFTLTFVIMILSLIPWEDFGIKFFTNINTWLTTMPILGGVIGKTMGAFGTWYF 316

Query: 332 PTMASSFLAITIIIMFLTATGANGIGERDVVDEFVNGASSLVGVSLIIGLARGINIILSQ 391
           P +   F+ + +++  +       + E D    F+ GA   +GV++I  +ARGI +I++
Sbjct: 317 PEITMLFIMMGVLVAIVYR-----MSEEDFFSSFLTGAGEFLGVAMICAIARGIQVIMNG 371

Query: 392 GYISDTMLYTASKLASHVSGSVFIIVMMFIYFVLGFVVPSSSGLAVLSMPILAPLADTVG 451
           G I+ T+L+      S +S  VF+I+    Y  + F++PS+SGLA  +M I+APL
Sbjct: 372 GMITATILHLGETSLSGLSSQVFVILAYIFYLPMSFLIPSTSGLAGATMGIMAPLGQFSN 431

Query: 452 IPRSVVVMAYQFGQYAMLFLAPT-GLVMATLQMLDMKYSHWLKFVWPVVLFLLIFGGGLLVL 512
           +P  +V+ A+Q     +  ++PT  +VM  L +  +    W KF+   ++ +++    LLV+
Sbjct: 432 VPAHLVITAFQSASGILNMISPTSAIVMGALALGRVDLGTWWKFIGKFIVMVMLVSVLLLVV 493
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1056

A DNA sequence (GBSx1129) was identified in *S.agalactiae* <SEQ ID 3255> which encodes the amino acid sequence <SEQ ID 3256>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.83   Transmembrane   25 -  41 (  18 -  47)
   INTEGRAL    Likelihood =-10.46   Transmembrane  153 - 169 ( 148 - 176)

----- Final Results -----
                 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13183 GB:Z99110 similar to two-component sensor histidine
           kinase [YkoG] [Bacillus subtilis]
 Identities = 119/446 (26%), Positives = 212/446 (46%), Gaps = 18/446 (4%)

Query: 17  TQITLWYSSFIFILVIGVLIGSFFISKSIAENKSKKNLEAKAVQMSQALAKGHRYEAFED 76
            T+I L+ S  + IL+I V     + I  S   +K    L  +   +++AL
Sbjct: 5   TKIHLYTSISLLILLILVHTAVYLIFSSALTSKDAARLADETDNIAEALRAAETEGVALQ 64

Query: 77  GIFYSVYDQNGKV-IYSGFPKGFKRDLDHQHKHKKKLSLFSMEN--------RTFQYVDI 127
             +  +    NG V + +G  K        +      LS  S E         + F    +
Sbjct: 65  DMLQAYLPANGMVRVVNGDQKAVMTITKEKAYKDFPLSFHSGETADVRKPDGKLFAEAAV 124

Query: 128 PISGKNQWLRAIRTVDRLDKQLTELLFSLGIVLPLMLIIITVG----GYLILKRTFRPIQ 183
            P+   +  + +++ V+RL+    E LF L I+L      + +     G L+ +R   PI+
Sbjct: 125 PVIWTDGQVVSLQLVERLENT-EESLFLLKIILIAASAAVCIASFFAGSLLARRIINPIR 183

Query: 184 EITETAQFITQNEDYTKRIITKNNENELTELAAVINTMLASIESSFVREKQFNNDVSHEL 243
             +  T + I +++++    +   + +EL ++    N M   ++  + +++QF  D SHEL
Sbjct: 184 RLMITMKDIQRDKEFKTISLEGQSNDELYQMGLTFNEMAMMLKEHYDKQQQFVQDASHEL 243

Query: 244 RTPVTVILSESEYGKNYAENLSEA-KESFEVIHRQSLSMKKLVEQLLELTKAENPLSIQL 302
            +TP+T+I S S   K +     E  +ES E IH +++ MKKL  QLL L K+   L + L
Sbjct: 244 KTPLTIIESYSSLMKRWGAKKPEVLEESIEAIHSEAVHMKKLTNQLLALAKSHQGLEVDL 303

Query: 303 EPLNFSIMMKQLVSDSSRLLDNTPIHLDSQIEDDLWIIGQQTLLKRLFDNLFSNAIKFTN 362
            + ++  I   + V  + + +    I L++  ++ L +   +  +K+L   L  NAIK++
Sbjct: 304 KTIDL-IKAARAVMQTLQSVYQRDILLETD-KESLLVKADEERIKQLLTILLDNAIKYSE 361

Query: 363 NHISISLRQSDNQIVFSIKDNGLGISVDDQSKIWNRFYQVDSARTKDSQSGIGLGLSLVK 422
              I +S    + +   S++D G+GI  +    ++ RFY+ D AR + +  G GLGLS+ K
Sbjct: 362 KPIEMSAGTRNGRPFLSVRDEGIGIPEEHIPHLFERFYRADEARNRKT-GGTGLGLSIAK 420

Query: 423 QIATIHRAKIWVDSKPDDGSQFTLTF 448
            QIA  H  ++ V SKP  G+  T+ F
Sbjct: 421 QIADEHGIELSVKSKPGQGTAVTMQF 446
```

There is also homology to SEQ ID 1178.

SEQ ID 3256 (GBS77) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 2; MW 78.5kDa) and in Figure 28 (lane 2; MW 78.5kDa).

GBS77-GST was purified as shown in Figure 195, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1057

A DNA sequence (GBSx1130) was identified in *S.agalactiae* <SEQ ID 3257> which encodes the amino acid sequence <SEQ ID 3258>. This protein is predicted to be CopR protein (tcrA). Analysis of this protein sequence reveals the following:

```
Possible site: 33
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3963(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC07978 GB:AJ278983 CopR protein [Ralstonia metallidurans]
 Identities = 102/221 (46%), Positives = 145/221 (65%)

Query: 1   MKILVVEDEFDLNRSIVKLLKKQHYSVDSASNGEEALQFVSVAEYDVIILDVMMPKMDGF 60
           MK+LVVEDE       + + L +  + VD  +NG +   F     YD+IILDVM+P +DG+
Sbjct: 1   MKLLVVEDEVKTGEYLRQGLTEAGFVVDLVANGLDGQHFAVNETYDLIILDVMLPDVDGW 60

Query: 61  TFLKLLRNKGSQVSILMLTARDAVEDRIAGLDFGADDYLVKPFEFGELMARIRAMLRRAN 120
             L  +R  G+ V +L LTARD+V DR+ GL+ GADDYLVKPF F EL+AR+R +LRR
Sbjct: 61  HILHAIRASGNAVPVLFLTARDSVADRVRGLELGADDYLVKPFAFSELLARVRTLLRRGA 120

Query: 121 RQVSSDDIQIQDITINLSTKQVWRNDNLIDLTAKEYEVLEYLARHRDQVLSRHQIREHVW 180
            Q++ D IQ+ D+ ++LS ++  R    I LT+KE+ +LE  AR R +VL R  I   VW
Sbjct: 121 VQLAMDRIQVADLILDLSRRRASRGGRRITLTSKEFALLELFARRRGEVLPRSLIASQVW 180

Query: 181 DYDYYGESNIIDVLIKNLRRKLDNNRDGSLIKTKRGLGYVI 221
           D ++  +SN+IDV I+ LR K+D+  +  LI+T RG+GYV+
Sbjct: 181 DMNFDSDSNVIDVAIRRLRAKIDDGFEVKLIQTVRGMGYVL 221
```

There is also homology to SEQ ID 3260.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1058

A DNA sequence (GBSx1131) was identified in *S.agalactiae* <SEQ ID 3261> which encodes the amino acid sequence <SEQ ID 3262>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.45   Transmembrane   18 -  34 (  16 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10281> which encodes amino acid sequence <SEQ ID 10282> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3262 (GBS78) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 4; MW 23.8kDa).

The GBS78-GST fusion product was purified (Figure 194, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 317), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1059

A DNA sequence (GBSx1132) was identified in *S.agalactiae* <SEQ ID 3263> which encodes the amino acid sequence <SEQ ID 3264>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.04   Transmembrane   15 -  31 (   6 -  35)
   INTEGRAL    Likelihood = -1.28   Transmembrane   51 -  67 (  51 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3264 (GBS79) was expressed in *E.coli* as a GST-fusion product. GBS79d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 17 & 18; MW 51kDa), in Figure 155 (lane 17; MW 51kDa) and in Figure 187 (lane 13; MW 51kDa). GBS79d was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 2-4; MW 26kDa) and in Figure 183 (lane 5; MW 26kDa). Purified GBS79d-GST is shown in Figure 243, lane 2.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1060

A DNA sequence (GBSx1133) was identified in *S.agalactiae* <SEQ ID 3265> which encodes the amino acid sequence <SEQ ID 3266>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5326(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10279> which encodes amino acid sequence <SEQ ID 10280> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG20974 GB:AE005164 Vng6349c [Halobacterium sp. NRC-1]
 Identities = 97/358 (27%), Positives = 163/358 (45%), Gaps = 20/358 (5%)

Query: 35  DPQIIKLTTRANIAIGTYEGFLESIINPMLLISPLLSQEAVLSSKLEGTHATLKDLLNYE 94
            D    +    A   +G   G    +  P +L + LL +EA+ S+++EG       L + E
Sbjct: 70  DDDFYETLADATFWLGKLSGVSLELDFPPVLYTSLLRKEAMESAEIEGADVDYDALYSLE 129

Query: 95  AGNKVDIERDELHEII------NYRKALFYALENISTINNIDSKGLPLSNRIIKEMHKIL 148
                 D  RDE  E        + R+ L Y       I+ +D+ G  L+  ++ ++H+ L
```

```
Sbjct: 130 T-RTFDEGRDEPSETTAAAETKDTREVLNYETAVKEGIDALDA-GEELNVELLHDLHETL 187

Query: 149 LDNV---RGSSKNPGNFKRSQNYIGSVSSISYTPVPAEKTPEYMSNLEQYIHYD-DLDLL 204
           L  V   R +    G++K + NY+G      + P       + M  L  Y         L
Sbjct: 188 LTGVPDDRVDTDTIGDYKTNPNYLGD-----FLPPAPGAVEDLMDGLFTYYRTGGSYHPL 242

Query: 205 VQSAIIHAQFEMIHPFEDGNGRIGRLLIPLFLYYQELLSYPTFYMSSYFERDRSLYISHL 264
           V  A+ H QFE IHP+ DGNGR+GRLLI L LY  +LL  P  Y+S Y  R+++ Y+  +
Sbjct: 243 VDIALFHYQFETIHPYGDGNGRLGRLLITLQLYDADLLERPNLYLSEYLNRNKTTYVERM 302

Query: 265 SNISKDNNWKDWFEYYLEGVILSAEESTKKAQDILSLYNIMKEQVIPKLNSVSGIQLLDF 324
             +     W+ W  +++EG+   A ES ++ + +  L    + +   K  + +  QL
Sbjct: 303 EGVRFHGEWEAWLSFFIEGIARQAHESVERTRALADLRREYEHEYGGKAYTKN--QLAVT 360

Query: 325 IFSAPIFKAEQVSEHLKISKRTTYTLLNKLIDEGYL-STDNAQRNRTYYCPQLLSIVQ 381
           +F  P   ++ V     I + T    +N+L++EG L       RN+ Y   ++  I++
Sbjct: 361 LFEQPYITSKTVQRLFDIEQSTASRAINELVNEGILEEVPRHGRNKEYRAREIFEILE 418
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1061

A DNA sequence (GBSx1134) was identified in *S.agalactiae* <SEQ ID 3267> which encodes the amino acid sequence <SEQ ID 3268>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4370(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif : 46-48
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3268 (GBS299) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 2; MW 62.2kDa) and in Figure 60 (lane 4; MW 62.2kDa).

GBS299-GST was purified as shown in Figure 207 (lane 4) and Figure 225 (lanes 2-3).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1062

A DNA sequence (GBSx1135) was identified in *S.agalactiae* <SEQ ID 3269> which encodes the amino acid sequence <SEQ ID 3270>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4176(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1063

A DNA sequence (GBSx1136) was identified in *S.agalactiae* <SEQ ID 3271> which encodes the amino acid sequence <SEQ ID 3272>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1789(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1064

A DNA sequence (GBSx1137) was identified in *S.agalactiae* <SEQ ID 3273> which encodes the amino acid sequence <SEQ ID 3274>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3748(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1065

A DNA sequence (GBSx1138) was identified in *S.agalactiae* <SEQ ID 3275> which encodes the amino acid sequence <SEQ ID 3276>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1638(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12294 GB:Z99106 similar to transposon protein [Bacillus subtilis]
 Identities = 84/291 (28%), Positives = 138/291 (46%), Gaps = 6/291 (2%)

Query: 6   MLDYLAVTIKGLAPDDVIEKILILPKDKFVLNEWGINKYQRHYSFSEIKVYFNKDWQSKM 65
           M+DY+ V+ K    D +IE++L L KD     + G   Y   Y    IKV+++    ++
Sbjct: 31  MVDYIRVSFKTHDVDRIIEEVLHLSKDFMTEKQSGFYGYVGTYELDYIKVFYSAPDDNR- 89

Query: 66  GVFIELRGQGCRQYEEYMENNVNNWVTLMKRISECHSNVTRLDIANDIFDDSLSVPLIYS 125
           GV IE+ GQGCRQ+E ++E     W    +   +   + TR D+A D      S+P +
Sbjct: 90  GVLIEMSGQGCRQFESFLECRKKTWYDFFQDCMQQGGSFTRFDLAIDDKKTYFSIPELLK 149

Query: 126 YCKKQLCISTAKTFDYHEKSLLENGEKVGEMVTIGVRGTQQW-CVYNKLLEQKLDQELPN 184
             +K  CIS  +  D++    L +G   G  +  G + ++ + C Y K  EQ     +P
Sbjct: 150 KAQKGECISRFRKSDFNGSFDLSDGITGGTTIYFGSKKSEAYLCFYEKNYEQAEKYNIPL 209

Query: 185 TPL-SWTRAELRCWQEKANLLAKQIKEGRPLKEIYFEVINGHYRFVSPRDKDSNRWRRKT 243
             L  W R ELR   E+A +    + + + L  I  ++IN + RFV   D++  R   KT
Sbjct: 210 EELGDWNRYELRLKNERAQVAIDALLKTKDLTLIAMQIINNYVRFVD-ADENITREHWKT 268

Query: 244 VKWWNDYLETQEKTVLSVKRTKPTLKRSEKWTEKQVSRTLGKLYVAKAESH 294
             +W+D++    +  L VK  K   ++S  W     + T+    V +A+ H
Sbjct: 269 SLFWSDFIGDVGRLPLYVKPQKDFYQKSRNWLRNSCAPTM--KMVLEADEH 317
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1066

A DNA sequence (GBSx1139) was identified in *S.agalactiae* <SEQ ID 3277> which encodes the amino acid sequence <SEQ ID 3278>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1914(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB70622 GB:AJ243106 integrase [Streptococcus thermophilus]
 Identities = 135/474 (28%), Positives = 233/474 (48%), Gaps = 68/474 (14%)

Query: 20  KAGNVLVKFAMRFTHPITKKSHKKYLSTGASKGWFTTKATPSKKLPSGKERLLVSDIKNT 79
           K G + VKF   F + +T K  ++ LS      W+T     +KK  +GK +L  S
Sbjct: 19  KTGYIEVKFRTYFNNQLTNK-RREILSD-----WYTIV---NKKDTTGKIKL--SPQIKA 67

Query: 80  QLITQVTQELNKLVDDYIAELMGIKPKKAKKLLTLEEIAKPFDKDGNFYGKAFKAWH--- 136
            +  ++ ++ NK+ ++    ++       K  +TL+E+              +  WH
Sbjct: 68  IIHKELQEKANKVYEELTRTIL-----LEKSDITLDEV--------------WNEWHNER 108

Query: 137 -ERVKPANNTLKTRVTIYNRYIEPNFDTRMSITKFAFMTDEIQNLIN-----ASSMHMAR 190
            ER   A  TL      Y  +I      + SI K    +  I+NL++      +   +A+
Sbjct: 109 VERQLVAPKTLAGEDGRYRNHITKQIP-KNSILK-NIPSSLIKNLLDNLYPIGNHKRLAQ 166

Query: 191 NLHIYLKMIFDWSVENGQITLTQDPIASNKVKRRVLTKSEEQDK-KREDIAEKYLEASEV 249
            +   L  I+ +++ +  I+  Q+P+    + R+ L  S+E D+ K+ DI ++YLE+ E+
Sbjct: 167 GVKSDLTSIYKFAILHDYISPDQNPMPYISIGRKGL--SDELDRLKKSDIEDQYLESWEL 224
```

```
Query: 250 NHVLRLIESWTNRPDNQLIADVLRMIFLTGMRPSEVLGLNEDMLDFEKKWIKVHWQRASK 309
             VL ++  +     N+  A +     LTGMR  EVLGL E+ +DF K    V   RA+
Sbjct: 225 KEVLSIVRKY-----NEQYARIFEFQALTGMRIGEVLGLKEEAIDFNKNIASVIRTRATH 279

Query: 310 NKSDDMMEALNLDEKERYRADLKTKESVRTIPMSPEVEKILRHYIDRNKFQAQFSPTYQD 369
             + +          + Y  ++K  +S R + +S    +IL+  I+ N    +F+P Y+D
Sbjct: 280 GGASE----------DSYEGNVKNLQSYRNVQLSKRAIEILKEEIELNHQHIRFNPDYKD 329

Query: 370 LGYLFTRTYIRAGNRQGSPLYHNELSQFLRGGSSQSAKYNKKAGKPYK---DIDSFLDFG 426
            G++FT   I   +  G+PL+++ L+ FL   SS++ K N+  G P +   DID+ L F
Sbjct: 330 NGWIFTSKSIHKPDYNGTPLHYSVLNNFL--NSSENGKLNRN-GNPRRAGIDIDNKLSFK 386

Query: 427 RPIHVIPHMFRHSFISIMASEGIDLPTIREFVGHSEDSKEIERVYLHVIKKQKD 480
           +  H+  H+FRH+ IS +A +G+ L  I++ VGHS  S+ +  +YLH+ KK KD
Sbjct: 387 K--HITTHIFRHTHISFLAEQGVPLEAIQDRVGHSRGSR-VTEIYLHITKKTKD 437
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3279> which encodes the amino acid sequence <SEQ ID 3280>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5203(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 82/357 (22%), Positives = 155/357 (42%), Gaps = 52/357 (14%)

Query: 135 WHERVKPANNTLKTRVTIYNRYIEPNFDTRMSITKFAFMTDEIQNLINA--SSMHMARNL 192
           W    K   +T    +    R +    D  + I K    T  +Q++I+    S    +
Sbjct: 73  WEHHQKSLKSTSVRSLDFRIRELRNLIDPEVMIAKIT--TKYLQSIIDKIPGSYDKRKRA 130

Query: 193 HIYLKMIFDWSVENGQITLTQDPIASNKVKRRVLTKSEEQDKKREDIAEKYLEASEVNHV 252
              LK  FD+++    +++  +P+ S ++ + V T      K  ED+A+K+LE  E+
Sbjct: 131 RQLLKQTFDYAIALEYVSI--NPVISTQLAKPVKTI-----KDFEDVAQKFLEKDELK-- 181

Query: 253 LRLIESWTNRPDNQLIADVLRMIFLTGMRPSEVLGLNEDMLDFEKKWIKVHWQRASKNKS 312
            RL++    R  +  +A +   + L G R  E L +  D  + +   I++H
Sbjct: 182 -RLLDEMYRRKGSIKMAYLAEFMSLNGCRIGEALAIQPD--NIKNDIIEIH--------- 229

Query: 313 DDMMEALNLDEKERYRADLKTKESVRTIPMSPEVEKILRHYIDRNKFQAQFSPTYQDLGY 372
              ++  +   +   +   KT  S R   ++   ++I++  +  N  +   +P Y+D+GY
Sbjct: 230 -GTLDYTSNGYRNAIKTTPKTNSSWRETLITKREKEIIQDILKINALEKNTNPNYKDMGY 288

Query: 373 LFTRTYIRAGNRQGSPLYHNELSQFLRGGSSQSAKYNKKAGKPYKDIDSFLDFGRPIHVI 432
           +F        +R G P+  N L+  +R         NK+  KP +             +
Sbjct: 289 IFI-------SRNGVPIQDNALNTSIRAA-------NKRLEKPIQK-----------ELT 323

Query: 433 PHMFRHSFISIMASEGIDLPTIREFVGHSEDSKEIERVYLHVIKKQKDTMRGAVEKL 489
            H+FRH+ +S +A   + L TI + VGH+ DSK  +++Y HV K  K+ +   + +L
Sbjct: 324 SHIFRHTLVSRLAENKVPLKTIMDRVGHA-DSKTTQQIYTHVTKSMKNEVVDILNRL 379
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1067

A DNA sequence (GBSx1140) was identified in *S.agalactiae* <SEQ ID 3281> which encodes the amino acid sequence <SEQ ID 3282>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3023(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10277> which encodes amino acid sequence <SEQ ID 10278> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB64982 GB:U43834 Ydr540cp [Saccharomyces cerevisiae]
 Identities = 88/170 (51%), Positives = 117/170 (68%), Gaps = 3/170 (1%)

Query: 36  MRTYSDKNELKEEVLKSYKKYIAEFNDIPEKLKDLRIDEVDRTPAENLAYQVGWTTLILK 95
            MR Y+ K ELKEE+ K Y+KY AEF  I E  KD +++ VDRTP+ENL+YQ+GW  L+L+
Sbjct: 1   MREYTSKKELKEEIEKKYEKYDAEFETISESQKDEKVETVDRTPSENLSYQLGWVNLLLE 60

Query: 96  WESDEQSGLEVKTPTETFKWNQLGELYQHFTETYASLTIKELTAQLNDNVDAIGNMIDSM 155
            WE+ E +G  V+TP   +KWN LG LYQ F + Y   +IKE  A+L + V+ +   I ++
Sbjct: 61  WEAKEIAGYNVETPAPGYKWNNLGGLYQSFYKKYGIYSIKEQRAKLREAVNEVYKWISTL 120

Query: 156 SDEVLFKPHMRNWADSATKNAVWEVYKFIHINTVAPFGTFRTKIRKWKKV 205
            SD+ LF+   R W   AT  A+W VYK+IHINTVAPF  FR KIRKWK++
Sbjct: 121 SDDELFQAGNRKW---ATTKAMWPVYKWIHINTVAPFTNFRGKIRKWKRL 167
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1068

A DNA sequence (GBSx1141) was identified in *S.agalactiae* <SEQ ID 3283> which encodes the amino acid sequence <SEQ ID 3284>. This protein is predicted to be 50S ribosomal protein subunit L33-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5420(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB66692 GB:U89998 50S ribosomal protein subunit L33
          [Lactococcus lactis subsp. cremoris]
 Identities = 43/49 (87%), Positives = 46/49 (93%)

Query: 1  MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHVVFTEVK 49
           MRVNITLEHKESGERLYLT KNKRNTPD+L+LKKYS KLRKHV+F EVK
Sbjct: 1  MRVNITLEHKESGERLYLTQKNKRNTPDKLELKKYSKKLRKHVIFKEVK 49
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3285> which encodes the amino acid sequence <SEQ ID 3286>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5394(Affirmative) < succ>
```

```
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 48/49 (97%), Positives = 48/49 (97%)

Query: 1  MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHVVFTEVK 49
           MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHV FTEVK
Sbjct: 1  MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHVTFTEVK 49
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1069

A DNA sequence (GBSx1142) was identified in *S.agalactiae* <SEQ ID 3287> which encodes the amino acid sequence <SEQ ID 3288>. This protein is predicted to be 50S ribosomal protein subunit L32-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3577(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB66691 GB:U89998 50S ribosomal protein subunit L32
          [Lactococcus lactis subsp. cremoris]
 Identities = 44/53 (83%), Positives = 48/53 (90%)

Query: 1  MAKPARHTSKAKRNKRRTHYKLTAPSVQFDETTGDYSRSHRVSLKGYYKGRKI 53
           MA PARHTS AK+N+RRTHYKLTAP+V FDETTGDY  SHRVSLKGYYKGRK+
Sbjct: 1  MAVPARHTSSAKKNRRRTHYKLTAPTVTFDETTGDYRHSHRVSLKGYYKGRKV 53
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3289> which encodes the amino acid sequence <SEQ ID 3290>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5148(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 38/39 (97%), Positives = 39/39 (99%)

Query: 22 LTAPSVQFDETTGDYSRSHRVSLKGYYKGRKIAKANEAK 60
           +TAPSVQFDETTGDYSRSHRVSLKGYYKGRKIAKANEAK
Sbjct: 1  MTAPSVQFDETTGDYSRSHRVSLKGYYKGRKIAKANEAK 39
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1070

A DNA sequence (GBSx1144) was identified in *S.agalactiae* <SEQ ID 3291> which encodes the amino acid sequence <SEQ ID 3292>. This protein is predicted to be histidyl-tRNA synthetase (hisS). Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4357(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10275> which encodes amino acid sequence <SEQ ID 10276> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA78919 GB:Z17214 histidine--tRNA ligase [Streptococcus
          equisimilis]
 Identities = 327/404 (80%), Positives = 361/404 (88%)

Query: 32  WQYVENVIRNLFKQYHYDEIRTPMFEHYEVISRSVGDTTDIVTKEMYDFHDKGDRHITLR 91
            WQYVE V R  FKQYHY EIRTPMFEHYEVISRSVGDTTDIVTKEMYDF+DKGDRHITLR
Sbjct: 1   WQYVEGVARETFKQYHYGEIRTPMFEHYEVISRSVGDTTDIVTKEMYDFYDKGDRHITLR 60

Query: 92  PEGTAPVVRSYVENKLFAPEVQKPTKMYYIGSMFRYERPQAGRLREFHQVGVECFGSNNP 151
            PEGTAPVVRSYVENKLFAPEVQKP K+YYIGSMFRYERPQAGRLREFHQ+GVECFGS NP
Sbjct: 61  PEGTAPVVRSYVENKLFAPEVQKPVKLYYIGSMFRYERPQAGRLREFHQIGVECFGSANP 120

Query: 152 ATDVETIAMGHHLFEDLGIKNVKLHLNSLGNPESRQAYRQALIDYLTPIREQLSKDSQRR 211
            ATDVETIAM +HLFE LGIK V LHLNSLGN  SR AYRQALIDYL+P+R+ LSKDSQRR
Sbjct: 121 ATDVETIAMAYHLFERLGIKGVTLHLNSLGNAASRAAYRQALIDYLSPMRDTLSKDSQRR 180

Query: 212 LNENPLRVLDSKEPEDKLAVENAPSILDYLDESSQAHFDAVCHMLDALNIPYIIDTNMVR 271
            L+ENPLRVLDSKE EDK+AV NAPSILDY DE SQAHFDAV  ML+AL IPY+IDTNMVR
Sbjct: 181 LDENPLRVLDSKEKEDKIAVANAPSILDYQDEESQAHFDAVRSMLEALAIPYVIDTNMVR 240

Query: 272 GLDYYNHTIFEFITEIEDNELTICAGGRYDGLVSYFGGPETPAFGFGLGLERLLLILDKQ 331
            GLDYYNHTIFEFITE++ +ELTICAGGRYDGLV YFGGP TP FGFGLGLERLLLILDKQ
Sbjct: 241 GLDYYNHTIFEFITEVDQSELTICAGGRYDGLVEYFGGPATPGFGFGLGLERLLLILDKQ 300

Query: 332 GISLPIENTIDLYIAVLGSEANLAALDLAQSIRHQGFKVERDYLGRKIKAQFKSADTFNA 391
            G+ LP+E  +D+YIAVLG++AN+AAL L Q+IR QGF VERDYLGRKIKAQFKSADTF A
Sbjct: 301 GVELPVEEGLDVYIAVLGADANVAALALTQAIRRQGFTVERDYLGRKIKAQFKSADTFKA 360

Query: 392 KVIMTLGSSEVDSKEVGLKNNQTRQEVKVSFENIKTDFSSVLKQ 435
            KV++TLG SE+ + +  LK+NQTRQE+ VSF+ I+TDF+S+  +
Sbjct: 361 KVVITLGESEIKAGQAVLKHNQTRQEMTVSFDQIQTDFASIFAE 404
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3293> which encodes the amino acid sequence <SEQ ID 3294>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3183(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 339/424 (79%), Positives = 387/424 (90%)
```

```
Query: 13  MKLQKPKGTQDILPGESAKWQYVENVIRNLFKQYHYDEIRTPMFEHYEVISRSVGDTTDI 72
           MKLQKPKGTQDILPG++AKWQYVE+V R+ F QY+Y EIRTPMFEHYEVISRSVGDTTDI
Sbjct: 1   MKLQKPKGTQDILPGDAAKWQYVESVARDTFSQYNYGEIRTPMFEHYEVISRSVGDTTDI 60

Query: 73  VTKEMYDFHDKGDRHITLRPEGTAPVVRSYVENKLFAPEVQKPTKMYYIGSMFRYERPQA 132
           VTKEMYDF+DKGDRHITLRPEGTAPVVRSYVENKLFAPEVQKP K+YYIGSMFRYERPQA
Sbjct: 61  VTKEMYDFYDKGDRHITLRPEGTAPVVRSYVENKLFAPEVQKPVKLYYIGSMFRYERPQA 120

Query: 133 GRLREFHQVGVECFGSNNPATDVETIAMGHHLFEDLGIKNVKLHLNSLGNPESRQAYRQA 192
           GRLREFHQ+GVECFG+ NPATDVETIAM +HLFE LGIK+V LHLNSLG+PESR AYRQA
Sbjct: 121 GRLREFHQIGVECFGAANPATDVETIAMAYHLFEKLGIKDVTLHLNSLGSPESRAAYRQA 180

Query: 193 LIDYLTPIREQLSKDSQRRLNENPLRVLDSKEPEDKLAVENAPSILDYLDESSQAHFDAV 252
           LIDYLTP+R+QLSKDSQRRL+ENPLRVLDSKE EDKLAVE APSILDYLDE SQAHF+AV
Sbjct: 181 LIDYLTPMRDQLSKDSQRRLDENPLRVLDSKEKEDKLAVEKAPSILDYLDEESQAHFEAV 240

Query: 253 CHMLDALNIPYIIDTNMVRGLDYYNHTIFEFITEIEDNELTICAGGRYDGLVSYFGGPET 312
             ML+AL+IPY+IDTNMVRGLDYY+HTIFEFIT +E ++LTICAGGRYD LV YFGGPET
Sbjct: 241 KDMLEALDIPYVIDTNMVRGLDYYSHTIFEFITSVEGSDLTICAGGRYDSLVGYFGGPET 300

Query: 313 PAFGFGLGLERLLLILDKQGISLPIENTIDLYIAVLGSEANLAALDLAQSIRHQGFKVER 372
           P FGFGLGLERLL+I++KQGI+LPIE  +D+Y+AVLG  AN  AL+L Q+IR QGF  ER
Sbjct: 301 PGFGFGLGLERLLMIIEKQGITLPIETEMDIYLAVLGDGANSKALELVQAIRRQGFTAER 360

Query: 373 DYLGRKIKAQFKSADTFNAKVIMTLGSSEVDSKEVGLKNNQTRQEVKVSFENIKTDFSSV 432
           DYLGRKIKAQFKSADTF AK++MTLG SEV++ +  +KNN++RQEV+VSFE++ T+F+++
Sbjct: 361 DYLGRKIKAQFKSADTFKAKLVMTLGESEVEAGKAVIKNNRSRQEVEVSFEDMMTNFANI 420

Query: 433 LKQL 436
            +QL
Sbjct: 421 SEQL 424
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1071

A DNA sequence (GBSx1145) was identified in *S.agalactiae* <SEQ ID 3295> which encodes the amino acid sequence <SEQ ID 3296>. This protein is predicted to be aspartyl-tRNA synthetase (aspS). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5124(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10273> which encodes amino acid sequence <SEQ ID 10274> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14714 GB:Z99118 aspartyl-tRNA synthetase [Bacillus subtilis]
 Identities = 339/585 (57%), Positives = 432/585 (72%), Gaps = 9/585 (1%)

Query: 20  RSMYAGRVRSEHIGTSITLKGWVGRRRDLGGLIFIDLRDREGIMQLVINPEEVSASVMAT 79
           R+ Y G +  + IG S+TLKGWV +RRDLGGLIFIDLRDR GI+Q+V NP+ VS   +A
Sbjct: 4   RTYYCGDITEKAIGESVTLKGWVQKRRDLGGLIFIDLRDRTGIVQVVFNPD-VSKEALAI 62

Query: 80  AESLRSEFVIEVSGVVTAREQA--NDNLPTGEVELKVQELSILNTSKTTPFEIKDGIE-A 136
           AE +R+E+V+++ G V ARE+   N NL TG +E+    +++LN +KT PF I D  E
```

```
Sbjct: 63  AEGIRNEYVLDIQGKVVAREEGTVNPNLKTGAIEIHADGVNVLNAAKTPPFAISDQAEEV 122

Query: 137 NDDTRMRYRYLDLRRPEMLENFKLRAKVTHSIRNYLDNLEFIDVETPMLTKSTPEGARDY 196
           ++D R+++RYLDLRRP M +  +LR  VT  ++R++LD   F+D+ETP+LT STPEGARDY
Sbjct: 123 SEDVRLKHRYLDLRRPAMFQTMQLRHNVTKAVRSFLDENGFLDIETPILTGSTPEGARDY 182

Query: 197 LVPSRVNQGHFYALPQSPQITKQLLMNAGFDRYYQIVKCFRDEDLRGDRQPEFTQVDLET 256
           LVPSRV++G FYALPQSPQ+ KQLLM +G +RYYQI +CFRDEDLR DRQPEFTQ+D+E
Sbjct: 183 LVPSRVHEGEFYALPQSPQLFKQLLMVSGIERYYQIARCFRDEDLRADRQPEFTQIDIEM 242

Query: 257 SFLSDQEIQDIVEGMIAKVMKDTKGLEVSLPFPRMAYDDAMNNYGSDKPDTRFDMLLQDL 316
           SF+S ++I  + E M+AKVM++TKG E+ LP PRM YD+AMN YGSDKPDTRFDMLL D+
Sbjct: 243 SFMSQEDIMSLAEEMMAKVMRETKGEELQLPLPRMTYDEAMNKYGSDKPDTRFDMLLTDV 302

Query: 317 TEIVKEVDFKVFSEA----SVVKAIVVKDKADKYSRKNIDKLTEIAKQYGAKGLAWLKYA 372
           ++IVK+ +FKVFS A     VVKAI VK  A  YSRK+ID L   A  YGAKGLAW+K
Sbjct: 303 SDIVKDTEFKVFSSAVANGGVVKAINVKGGAGDYSRKDIDALGAFAANYGAKGLAWVKVE 362

Query: 373 DNTISGPVAKFL-TAIEGRLTEALQLENNDLILFVADSLEVANETLGALRTRIAKELELI 431
            + + GP+AKF     + +L EAL     DL+LF AD  EV   +LGALR ++ KE  LI
Sbjct: 363 ADGVKGPIAKFFDEEKQSKLIEALDAAEGDLLLFGADQFEVVAASLGALRLKLGKERGLI 422

Query: 432 DYSKFNFLWVVDWPMFEWSEEEGRYMSAHHPFTLPTAETAHELEGDLAKVRAVAYDIVLN 491
           D   FNFLWV+DWP+ E   EEGR+ +AHHPFT+P  E    +E     ++A AYD+VLN
Sbjct: 423 DEKLFNFLWVIDWPLLEHDPEEGRFYAAHHPFTMPVREDLELIETAPEDMKAQAYDLVLN 482

Query: 492 GYELGGGSLRINQKDTQERMFKALGFSAESAQEQFGFLLEAMDYGFPPHGGLAIGLDRFV 551
           GYELGGGS+RI +KD QE+MF  LGFS E A EQFGFLLEA +YG PPHGG+A+GLDR V
Sbjct: 483 GYELGGGSIRIFEKDIQEKMFALLGFSPEEAAEQFGFLLEAFEYGAPPHGGIALGLDRLV 542

Query: 552 MLLAGKDNIREVIAFPKNNKASDPMTQAPSLVSEQQLEELSLTVE 596
           MLLAG+ N+R+ IAFPK   AS  MT+AP  VS+ QL+EL L+++
Sbjct: 543 MLLAGRTNLRDTIAFPKTASASCLMTEAPGEVSDAQLDELHLSIK 587
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3297> which encodes the amino acid sequence <SEQ ID 3298>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 495/582 (85%), Positives = 538/582 (92%)

Query: 18  MKRSMYAGRVRSEHIGTSITLKGWVGRRRDLGGLIFIDLRDREGIMQLVINPEEVSASVM 77
           MKRSMYAGRVR EHIGT+ITLKGWV RRRDLGGLIFIDLRDREG+MQLVINPEEVS+ VM
Sbjct: 18  MKRSMYAGRVREEHIGTTITLKGWVSRRRDLGGLIFIDLRDREGVMQLVINPEEVSSDVM 77

Query: 78  ATAESLRSEFVIEVSGVVTAREQANDNLPTGEVELKVQELSILNTSKTTPFEIKDGIEAN 137
           ATAE LRSE+VIEV G V AR+QAND L TG VELKV  L+ILNT+KTTPFEIKD +E +
Sbjct: 78  ATAERLRSEYVIEVEGFVEARQQANDKLATGMVELKVSALTILNTAKTTPFEIKDDVEVS 137

Query: 138 DDTRMRYRYLDLRRPEMLENFKLRAKVTHSIRNYLDNLEFIDVETPMLTKSTPEGARDYL 197
           DDTR+RYRYLDLRRPEMLENFKLRAKVTHSIRNYLD+LEFIDVETPMLTKSTPEGARDYL
Sbjct: 138 DDTRLRYRYLDLRRPEMLENFKLRAKVTHSIRNYLDDLEFIDVETPMLTKSTPEGARDYL 197

Query: 198 VPSRVNQGHFYALPQSPQITKQLLMNAGFDRYYQIVKCFRDEDLRGDRQPEFTQVDLETS 257
           VPSRV+QGHFYALPQSPQITKQLLMNAGFDRYYQIVKCFRDEDLRGDRQPEFTQVDLETS
Sbjct: 198 VPSRVSQGHFYALPQSPQITKQLLMNAGFDRYYQIVKCFRDEDLRGDRQPEFTQVDLETS 257

Query: 258 FLSDQEIQDIVEGMIAKVMKDTKGLEVSLPFPRMAYDDAMNNYGSDKPDTRFDMLLQDLT 317
           FLS+QEIQDIVEGMIAKVMK+TK ++V+LPFPRM+YD AMN+YGSDKPDTRF+MLLQDLT
```

```
Sbjct: 258 FLSEQEIQDIVEGMIAKVMKETKEIDVTLPFPRMSYDVAMNSYGSDKPDTRFEMLLQDLT 317

Query: 318 EIVKEVDFKVFSEASVVKAIVVKDKADKYSRKNIDKLTEIAKQYGAKGLAWLKYADNTIS 377
              VK  DFKVFSEA  VKAIVVK  AD+YSRK+IDKLTE AKQ+GAKGLAW+K  D  ++
Sbjct: 318 VTVKGNDFKVFSEAPAVKAIVVKGNADRYSRKDIDKLTEFAKQFGAKGLAWVKVTDGQLA 377

Query: 378 GPVAKFLTAIEGRLTEALQLENNDLILFVADSLEVANETLGALRTRIAKELELIDYSKFN 437
            GPVAKFLTAIE  L+  L+L  NDL+LFVAD+LEVAN TLGALR RIAK+L++ID S+FN
Sbjct: 378 GPVAKFLTAIETELSSQLKLAENDLVLFVADTLEVANNTLGALRNRIAKDLDMIDQSQFN 437

Query: 438 FLWVVDWPMFEWSEEEGRYMSAHHPFTLPTAETAHELEGDLAKVRAVAYDIVLNGYELGG 497
            FLWVVDWPMFEWSEEEGRYMSAHHPFTLPT E+AHELEGDLAKVRA+AYDIVLNGYELGG
Sbjct: 438 FLWVVDWPMFEWSEEEGRYMSAHHPFTLPTPESAHELEGDLAKVRAIAYDIVLNGYELGG 497

Query: 498 GSLRINQKDTQERMFKALGFSAESAQEQFGFLLEAMDYGFPPHGGLAIGLDRFVMLLAGK 557
            GSLRINQK+ QERMFKALGF+A+ A +QFGFLLEAMDYGFPPHGGLAIGLDRFVMLLAGK
Sbjct: 498 GSLRINQKEMQERMFKALGFTADEANDQFGFLLEAMDYGFPPHGGLAIGLDRFVMLLAGK 557

Query: 558 DNIREVIAFPKNNKASDPMTQAPSLVSEQQLEELSLTVESYE 599
            DNIREVIAFPKNNKASDPMTQAPSLVSE QLEELSL +ES++
Sbjct: 558 DNIREVIAFPKNNKASDPMTQAPSLVSENQLEELSLQIESHD 599
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1072

A DNA sequence (GBSx1146) was identified in *S.agalactiae* <SEQ ID 3299> which encodes the amino acid sequence <SEQ ID 3300>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.44   Transmembrane  186 - 202 ( 182 - 205)
   INTEGRAL    Likelihood = -5.68   Transmembrane   88 - 104 (  86 - 106)
   INTEGRAL    Likelihood = -3.40   Transmembrane  115 - 131 ( 112 - 132)
   INTEGRAL    Likelihood = -2.13   Transmembrane  141 - 157 ( 141 - 157)
   INTEGRAL    Likelihood = -0.96   Transmembrane   43 -  59 (  43 -  59)

----- Final Results -----
            bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12952 GB:Z99109 alternate gene name: yuxA~similar to
           hypothetical proteins [Bacillus subtilis]
 Identities = 104/275 (37%), Positives = 181/275 (65%), Gaps = 1/275 (0%)

Query: 39  EKISASLLYGILSSVAVNFFFQPGHVYSSGATGLAQVISAVSKHWFSFEIPVALAFYAIN 98
            +K+   ++  +L++  +N F  P  VY+SG TG+AQ++S+V   +  F I     +  +N
Sbjct: 7   KKLLIVIIGALLNAAGLNLFLIPADVYASGFTGVAQLLSSVVDQYAPFYISTGTLLFLLN 66

Query: 99  IPLLILSWRKIGHKFTIFTFITVTVSSIFIQLMPQITLTTDPLINAIFGGLIMGAGVGFS 158
            IP+ IL W K+G  FT+++ ++V ++++F+ ++P+ +L+ D L+NA+FGG+I   G+G +
Sbjct: 67  IPVGILGWLKVGKSFTVYSILSVALTTLFMGILPETSLSHDILLNAVFGGVISAVGIGLT 126

Query: 159 FKSRISSGGTDIISLTIRKKTGRDVGSISFIINGIILLFAGLLFGWKYALYSMVTIFVSS 218
             K   S+GG DI+++ + K   + VG+  FI+NGII+L AGLL GW+ ALY++VT++V++
Sbjct: 127 LKYGASTGGLDIVAMVLAKWKDKPVGTYFFILNGIIILTAGLLQGWEKALYTLVTLYVTT 186

Query: 219 RVTDAIFTKQKKMQAMIVTSKPYCVIKRIHRDLHRGVTCINDAEGTYNHEKKAVLITILT 278
            RV DAI T+  K+ AMIVT K   + + I+  + RG+T +  A+G + +E+K ++I ++T
Sbjct: 187 RVIDAIHTRHMKLTAMIVTKKADEIKEAIYGKMVRGITTV-PAKGAFTNEQKEMMIIVIT 245

Query: 279 REEFSDFKYLMLKADPKAFVSVAENVHIIGRFVDD 313
```

```
           R E  D + ++ + DPKAF ++ +   I G F  D
Sbjct: 246 RYELYDLEKIVKEVDPKAFTNIVQTTGIFGFFRKD 280
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3301> which encodes the amino acid sequence <SEQ ID 3302>. Analysis of this protein sequence reveals the following:

```
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.47   Transmembrane   87 - 103 (  86 - 106)
   INTEGRAL    Likelihood = -4.94   Transmembrane  185 - 201 ( 182 - 203)
   INTEGRAL    Likelihood = -1.59   Transmembrane  114 - 130 ( 113 - 130)
   INTEGRAL    Likelihood = -1.12   Transmembrane   42 -  58 (  42 -  58)
   INTEGRAL    Likelihood = -0.32   Transmembrane  140 - 156 ( 140 - 156)

----- Final Results -----
               bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA66894 GB:X98238 orf2 [Lactobacillus sakei]
 Identities = 105/280 (37%), Positives = 180/280 (63%), Gaps = 7/280 (2%)

Query: 37  AEKISASLLYGILSSIAVNFFFQPGHVYSSGATGLAQVFSAL-SHRLLGYDFPIAFAFYL 95
            +++I  +++YG L++++VN F  P   YSSG TG+AQ+ +AL SH  LG    +A  ++
Sbjct: 8   SKRIVIAMVYGFLAAVSVNLFLIPAKTYSSGVTGVAQLLTALVSH--LGGSLSVAALVFI 65

Query: 96  INIPLLILAWYKIGHQFTIFTFITVSMSSFFIQIMPQVT--LTTDPLINAIFGGLVMGMG 153
            +N+PLL+LAW+KI HQ+ IF+ + V  S  F++I+P     + T+    A+FGG ++G+G
Sbjct: 66  LNVPLLVLAWFKINHQYAIFSIVAVFTSVIFLKIIPVPVQPILTERFAGALFGGALIGLG 125

Query: 154 IGTGLKSRISSGGTDIVSLTLRKRTGKDVGSLSLMVNGAILAFAGILFGWQYALYSMVSI 213
            +G   ++  S+GGTD++   + + TGK VG+++ ++NG I+  AGI FGW  ALYS+V I
Sbjct: 126 VGLCFRAGFSTGGTDVIVTLVGRLTGKRVGAVNNVINGMIILAAGIFFGWGAALYSIVEI 185

Query: 214 FVSSRVTDAIFTKQKKMQATIVTSHPERVIHMIHKRLHRGVTSINDAEGTYKHEQKAVLI 273
            FVSS + D I+T+Q+K+  TI T  PE +   + + +H G T + D  G Y +++ +V++
Sbjct: 186 FVSSLLMDYIYTQQQKVTVTIFTKQPEALKKRMREFIH-GATEL-DGTGLYTNQETSVIM 243

Query: 274 TILTCEEYPEFKWLMLKTDPQAFVSVAENVRIIGRFVEDD 313
            T+++  +    K ++   DP AFV++   + + GRF  ++
Sbjct: 244 TVVSKYDLTALKLVVQDADPNAFVNIQSTMNLWGRFESNE 283
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 239/311 (76%), Positives = 274/311 (87%)

Query: 4   RRTPLEKKVKYIISVWAKKFGLLHTLKSISREKYAEKISASLLYGILSSVAVNFFFQPGH 63
            ++T  +KKVKY+IS  AKK GLLH L+SISREKYAEKISASLLYGILSS+AVNFFFQPGH
Sbjct: 3   KKTTYKKKVKYVISRGAKKVGLLHALRSISREKYAEKISASLLYGILSSIAVNFFFQPGH 62

Query: 64  VYSSGATGLAQVISAVSKHWFSFEIPVALAFYAINIPLLILSWRKIGHKFTIFTFITVTV 123
            VYSSGATGLAQV SA+S     ++ P+A AFY INIPLLIL+W KIGH+FTIFTFITV++
Sbjct: 63  VYSSGATGLAQVFSALSHRLLGYDFPIAFAFYLINIPLLILAWYKIGHQFTIFTFITVSM 122

Query: 124 SSIFIQLMPQITLTTDPLINAIFGGLIMGAGVGFSFKSRISSGGTDIISLTIRKKTGRDV 183
            SS FIQ+MPQ+TLTTDPLINAIFGGL+MG G+G   KSRISSGGTDI+SLT+RK+TG+DV
Sbjct: 123 SSFFIQIMPQVTLTTDPLINAIFGGLVMGMGIGTGLKSRISSGGTDIVSLTLRKRTGKDV 182

Query: 184 GSISFIINGIILLFAGLLFGWKYALYSMVTIFVSSRVTDAIFTKQKKMQAMIVTSKPYCV 243
            GS+S ++NG IL FAG+LFGW+YALYSMV+IFVSSRVTDAIFTKQKKMQA IVTS P  V
Sbjct: 183 GSLSLMVNGAILAFAGILFGWQYALYSMVSIFVSSRVTDAIFTKQKKMQATIVTSHPERV 242

Query: 244 IKRIHRDLHRGVTCINDAEGTYNHEKKAVLITILTREEFSDFKYLMLKADPKAFVSVAEN 303
            I  IH+ LHRGVT INDAEGTY HE+KAVLITILT EE+ +FK+LMLK DP+AFVSVAEN
Sbjct: 243 IHMIHKRLHRGVTSINDAEGTYKHEQKAVLITILTCEEYPEFKWLMLKTDPQAFVSVAEN 302
```

```
Query: 304 VHIIGRFVDDD 314
           V IIGRFV+DD
Sbjct: 303 VRIIGRFVEDD 313
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1073

A DNA sequence (GBSx1147) was identified in *S.agalactiae* <SEQ ID 3303> which encodes the amino acid sequence <SEQ ID 3304>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.72   Transmembrane  156 - 172 ( 156 - 174)
   INTEGRAL    Likelihood = -3.03   Transmembrane  112 - 128 ( 110 - 129)
   INTEGRAL    Likelihood = -2.34   Transmembrane   80 -  96 (  79 -  96)
   INTEGRAL    Likelihood = -1.49   Transmembrane   60 -  76 (  58 -  76)

----- Final Results -----
              bacterial membrane --- Certainty=0.2487(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05397 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 113/278 (40%), Positives = 192/278 (68%), Gaps = 1/278 (0%)

Query: 7   KTKIKETILIAFGVALYTFGFVKFNMANHLAEGGISGVTLIIHALFGVNPALSSLLLNIP 66
           + K K  + I  G A+++FG V FNM N+LAEGG +G+TLI++ +F +NPA+++L+LNIP
Sbjct: 4   RLKWKNIVFILLGSAIFSFGLVYFNMENNLAEGGFTGITLILYFMFQINPAVTNLVLNIP 63

Query: 67  LFILGARILGKKSLLLTIYGTVLMSFFMWFWQQIP-VTVPLKNDMMLVAVAAGILAGTGS 125
           + ++G +ILG+ +L+ TI GTV +S F+  +Q+   + +PL +DM L A+ AG+  GTG
Sbjct: 64  ILLIGWKILGRVTLIYTIIGTVSVSVFLEMFQRWKFMDIPLHDDMTLAALFAGVFVGTGL 123

Query: 126 GLVFRYGATTGGADIIGRIVEEKSGIKLGQTLLFIDAIVLTSSLVYINLQQMLYTLVASF 185
           G+VFR+G TTGG DII ++     G  +G+T+   DA+V+ SSL+Y+N ++ +YTL+A F
Sbjct: 124 GIVFRFGGTTGGVDIIAKLGFRYLGWSMGKTMFMFDAVVIASSLIYLNYREAMYTLLAVF 183

Query: 186 VFSQVLTNVENGGYTVRGMIIITKESESAAATILHEINRGVTFLRGQGAYSGREHDVLYV 245
           + ++V+  ++   Y+ +   II++ +E+ A TIL E+ RG T L+G+G+++G E ++LY
Sbjct: 184 IAAKVIDFIQQTAYSAKAAFIISEHTEAIADTILKEMERGATTLKGKGSFTGTEKEILYC 243

Query: 246 ALNPSEVRDVKEIMADLDPDAFISVINVDEVISSDFKI 283
            +  +E+  +K ++  +DP AF++V +V +VI   F +
Sbjct: 244 VVGRNELIRLKSLVERIDPHAFVTVNDVQDVIGEGFTL 281
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3305> which encodes the amino acid sequence <SEQ ID 3306>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.15   Transmembrane  112 - 128 ( 109 - 130)
   INTEGRAL    Likelihood = -2.34   Transmembrane  156 - 172 ( 156 - 174)
   INTEGRAL    Likelihood = -1.81   Transmembrane  178 - 194 ( 177 - 194)
   INTEGRAL    Likelihood = -1.65   Transmembrane   80 -  96 (  79 -  96)
   INTEGRAL    Likelihood = -0.37   Transmembrane   60 -  76 (  59 -  76)

----- Final Results -----
              bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05397 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 116/276 (42%), Positives = 182/276 (65%), Gaps = 1/276 (0%)

Query: 9   KLLKLFLIALGVAIYTFGFVNFNMANALAEGGVAGITLILHAHFGINPAYSSLLFNLPLF 68
            K   +  I LG AI++FG V FNM N LAEGG  GITLIL+  F INPA ++L+ N+P+
Sbjct: 6   KWKNIVFILLGSAIFSFGLVYFNMENNLAEGGFTGITLILYFMFQINPAVTNLVLNIPIL 65

Query: 69  ILGAKIFGKRSLALTIYGTVLMSAFIWMWQKVP-IELGLENDMMLVAVVAGLFSGIGSGI 127
            ++G KI G+ +L  TI GTV +S F+ M+Q+   +++ L +DM L A+ AG+F G G GI
Sbjct: 66  LIGWKILGRVTLIYTIIGTVSVSVFLEMFQRWKFMDIPLHDDMTLAALFAGVFVGTGLGI 125

Query: 128 VFRYGATTGGTDIIGRIAEEKFGAKLGQTLLLVDALVLTASLTYVDLKHMLYTLVASFVF 187
            VFR+G TTGG DII ++      G  +G+T+ + DA+V+ +SL Y++ +  +YTL+A F+
Sbjct: 126 VFRFGGTTGGVDIIAKLGFRYLGWSMGKTMFMFDAVVIASSLIYLNYREAMYTLLAVFIA 185

Query: 188 SQMISVVQNGGYTIRGMIIITKHSEAAAQAILTEINRGVTYLKGQGAYSGNDYNIMYVTL 247
            +++I  +Q   Y+ +   II++H+EA A  IL E+ RG T LKG+G+++G +  I+Y  +
Sbjct: 186 AKVIDFIQQTAYSAKAAFIISEHTEAIADTILKEMERGATTLKGKGSFTGTEKEILYCVV 245

Query: 248 NPTEVREVKRILAGLDPDAFISIIDVDEVISSDFKI 283
               E+  +K ++  +DP AF+++ DV +VI   F +
Sbjct: 246 GRNELIRLKSLVERIDPHAFVTVNDVQDVIGEGFTL 281
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 206/286 (72%), Positives = 250/286 (87%)

Query: 5   DLKTKIKETILIAFGVALYTFGFVKFNMANHLAEGGISGVTLIIHALFGVNPALSSLLLN 64
            D  TK+ +  LIA GVA+YTFGFV FNMAN LAEGG++G+TLI+HA FG+NPA SSLL N
Sbjct: 5   DKLTKLLKLFLIALGVAIYTFGFVNFNMANALAEGGVAGITLILHAHFGINPAYSSLLFN 64

Query: 65  IPLFILGARILGKKSLLLTIYGTVLMSFFMWFWQQIPVTVPLKNDMMLVAVAAGILAGTG 124
            +PLFILGA+I GK+SL LTIYGTVLMS F+W WQ++P+ + L+NDMMLVAV AG+ +G G
Sbjct: 65  LPLFILGAKIFGKRSLALTIYGTVLMSAFIWMWQKVPIELGLENDMMLVAVVAGLFSGIG 124

Query: 125 SGLVFRYGATTGGADIIGRIVEEKSGIKLGQTLLFIDAIVLTSSLVYINLQQMLYTLVAS 184
            SG+VFRYGATTGG DIIGRI EEK G KLGQTLL +DA+VLT+SL Y++L+ MLYTLVAS
Sbjct: 125 SGIVFRYGATTGGTDIIGRIAEEKFGAKLGQTLLLVDALVLTASLTYVDLKHMLYTLVAS 184

Query: 185 FVFSQVLTNVENGGYTVRGMIIITKESESAAATILHEINRGVTFLRGQGAYSGREHDVLY 244
            FVFSQ+++ V+NGGYT+RGMIIITK SE+AA  IL EINRGVT+L+GQGAYSG +++++Y
Sbjct: 185 FVFSQMISVVQNGGYTIRGMIIITKHSEAAAQAILTEINRGVTYLKGQGAYSGNDYNIMY 244

Query: 245 VALNPSEVRDVKEIMADLDPDAFISVINVDEVISSDFKIRRRNYDK 290
            V LNP+EVR+VK I+A LDPDAFIS+I+VDEVISSDFKIRRRNYDK
Sbjct: 245 VTLNPTEVREVKRILAGLDPDAFISIIDVDEVISSDFKIRRRNYDK 290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1074

A DNA sequence (GBSx1148) was identified in *S.agalactiae* <SEQ ID 3307> which encodes the amino acid sequence <SEQ ID 3308>. This protein is predicted to be BacB protein. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4355(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11330 GB:D78257 BacB [Enterococcus faecalis]
 Identities = 27/88 (30%), Positives = 48/88 (53%), Gaps = 1/88 (1%)

Query: 1  MPSEKEILDALSKVYSEEVIQADDYFRQAIFELASQLEKEGMN-SLLATKIDSLINQYVL 59
          M  ++E+LD LSK Y++  I   +  +  +FE A +L     N   +  K+ ++ ++Y+
Sbjct: 1  MDKQQELLDLLSKAYNDPKINEYEGLKDKLFECAKRLTTNETNIGEVCYKLSTINSEYLA 60

Query: 60 THQFDAPKSIFDLSRLVKTKASHYKGTA 87
           H F+ PKSI +L + V  +   Y+G A
Sbjct: 61 RHHFEMPKSIIELQKFVTKEGQKYRGWA 88
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3309> which encodes the amino acid sequence <SEQ ID 3310>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2712(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 99/102 (97%), Positives = 102/102 (99%)

Query: 1   MPSEKEILDALSKVYSEEVIQADDYFRQAIFELASQLEKEGMNSLLATKIDSLINQYVLT 60
           MPSEKEILDALSKVYSE+VIQADDYFRQAIFELASQLEKEGM+SLLATKIDSLINQY+LT
Sbjct: 7   MPSEKEILDALSKVYSEQVIQADDYFRQAIFELASQLEKEGMSSLLATKIDSLINQYILT 66

Query: 61  HQFDAPKSIFDLSRLVKTKASHYKGTAISAIMLGSFLSGGPK 102
           HQFDAPKSIFDLSRLVKTKASHYKGTAISAIMLGSFLSGGPK
Sbjct: 67  HQFDAPKSIFDLSRLVKTKASHYKGTAISAIMLGSFLSGGPK 108
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1075

A DNA sequence (GBSx1149) was identified in *S.agalactiae* <SEQ ID 3311> which encodes the amino acid sequence <SEQ ID 3312>. This protein is predicted to be ArgS (argS). Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2522(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10271> which encodes amino acid sequence <SEQ ID 10272> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF86984 GB:AF282249 ArgS [Lactococcus lactis subsp. lactis]
 Identities = 377/566 (66%), Positives = 464/566 (81%), Gaps = 5/566 (0%)
```

```
Query: 12  MDTKHLIASEIQKVVPD-MEQSTILSLLETPKNSSMGDLAFPAFSLAKTLRKAPQIIASD 70
           MD K L++  +   +   +    I +++E PK+S +GDLAFPAF LAKTLRK+PQIIA +
Sbjct: 1   MDEKQLVSQALSAAIDGVLGVEQIAAIIEKPKSSDLGDLAFPAFQLAKTLRKSPQIIAGE 60

Query: 71  IAEQIKSDQFEKVEAVGPYVNFFLDKAAISSQVLKQVLSDGSAYATQNIGEGRNVAIDMS 130
           IAE+I +  FEKV AVGPYVNFFLDK A +S+V+++VL++G  Y   NIGEG NV IDMS
Sbjct: 61  IAEKIDTKGFEKVIAVGPYVNFFLDKNATASEVIREVLAEGEHYGDANIGEGGNVPIDMS 120

Query: 131 SPNIAKPFSIGHLRSTVIGDSLANIFDKIGYHPVKINHLGDWGKQFGMLIVAYKKWGNEE 190
           +PNIAKPFSIGHLRSTVIGDS+A I++K+GY P+KINHLGDWGKQFG+LI AYKK+G+E
Sbjct: 121 APNIAKPFSIGHLRSTVIGDSIAKIYEKLGYQPIKINHLGDWGKQFGLLITAYKKYGDEA 180

Query: 191 AVRAHPIDELLKLYVRINAEAETDPSVDEEAREWFRKLEANDPEATELWQWFRDESLLEF 250
            + A+PIDELLKLYV+INAEA+ D  VDEE R+WF K+E  D EA  +W+WF D SL+EF
Sbjct: 181 TITANPIDELLKLYVKINAEAKEDSEVDEEGRQWFLKMEQGDEEALRIWKWFSDVSLIEF 240

Query: 251 NRLYDQMNVTFDSYNGEAFYNDKMDEVLELLESKNLLVESKGAQVVNLEKYGIEHPALIK 310
           NR+Y ++ VTFD + GE+FY+DKMD ++E LE+KNLL ESKGA +V+LEKY + +PALIK
Sbjct: 241 NRIYGKLGVTFDHFMGESFYSDKMDAIVEDLENKNLLHESKGALIVDLEKYNL-NPALIK 299

Query: 311 KSDGATLYITRDLAAALYRKRTYDFAKSIYVVGNEQSAHFKQLKAVLKEMDYDWSDDMTH 370
           K+DGATLYITRDLA A YRK+T++F KS+YVVG EQ+ HFKQLKAVLKE  YDWSDDM H
Sbjct: 300 KTDGATLYITRDLATAAYRKKTFNFVKSLYVVGGEQTNHFKQLKAVLKEAGYDWSDDMVH 359

Query: 371 VPFGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADKDKVAQAVGVGAI 430
           VPFG+VT+GG K STRKG+V+ LE  + EA++RA  QIEAKNPNL +K++VA+ VGVGA+
Sbjct: 360 VPFGMVTQGGKKFSTRKGHVVKLEMALDEAVDRAEKQIEAKNPNLENKEEVAKQVGVGAV 419

Query: 431 KFYDLKTDRTNGYDFDLEAMVSFEGETGPYVQYAHARIQSILRKANFSPSNSDNYSL--N 488
           KFYDLKTDR NGYDFDL+ MVSFEGETGPYVQYAHARIQSILRKAN    N DN SL  +
Sbjct: 420 KFYDLKTDRNNGYDFDLDEMVSFEGETGPYVQYAHARIQSILRKAN-RKVNIDNISLVVS 478

Query: 489 DVESWEIIKLIQDFPRIIVRAADNFEPSIIAKFAINLAQCFNKYYAHTRILDEDAEISSR 548
           D E+WEI+K +++FP I+ RAADN+EPSIIAK+AI+LAQ FNKYYAH RIL++DA++  R
Sbjct: 479 DAEAWEIVKALKEFPNIVKRAADNYEPSIIAKYAISLAQAFNKYYAHVRILEDDAQLDGR 538

Query: 549 LALCYATATVLKESLRLLGVDAPNEM 574
           LAL  AT+ VLKE+LRLLGV AP  M
Sbjct: 539 LALISATSIVLKEALRLLGVAAPENM 564
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3313> which encodes the amino acid sequence <SEQ ID 3314>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1734(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 492/563 (87%), Positives = 526/563 (93%)

Query: 12  MDTKHLIASEIQKVVPDMEQSTILSLLETPKNSSMGDLAFPAFSLAKTLRKAPQIIASDI 71
           MDTK LIASEI KVVP++EQ  I +LLETPKNS MGDLAFPAFSLAK LRKAPQ+IAS++
Sbjct: 1   MDTKTLIASEIAKVVPELEQDAIFNLLETPKNSDMGDLAFPAFSLAKVLRKAPQMIASEL 60

Query: 72  AEQIKSDQFEKVEAVGPYVNFFLDKAAISSQVLKQVLSDGSAYATQNIGEGRNVAIDMSS 131
           AEQI   QFEKV AVGPY+NFFLDKA ISSQVL+QV++ GS YA Q+ G+GRNVAIDMSS
Sbjct: 61  AEQIDESQFEKVVAVGPYINFFLDKAKISSQVLEQVITAGSDYAQQDEGQGRNVAIDMSS 120

Query: 132 PNIAKPFSIGHLRSTVIGDSLANIFDKIGYHPVKINHLGDWGKQFGMLIVAYKKWGNEEA 191
           PNIAKPFSIGHLRSTVIGDSLA+IF K+GY PVKINHLGDWGKQFGMLIVAYKKWG+E A
Sbjct: 121 PNIAKPFSIGHLRSTVIGDSLAHIFAKMGYKPVKINHLGDWGKQFGMLIVAYKKWGDEAA 180
```

```
Query: 192 VRAHPIDELLKLYVRINAEAETDPSVDEEAREWFRKLEANDPEATELWQWFRDESLLEFN 251
           V+AHPIDELLKLYVRINAEAETDP+VDEEAREWFRKLE  D EATELWQWFRDESLLEFN
Sbjct: 181 VQAHPIDELLKLYVRINAEAETDPTVDEEAREWFRKLEDGDKEATELWQWFRDESLLEFN 240

Query: 252 RLYDQMNVTFDSYNGEAFYNDKMDEVLELLESKNLLVESKGAQVVNLEKYGIEHPALIKK 311
           RLYDQ++VTFDSYNGEAFYNDKMDEVL+LLE+KNLLVESKGAQVVNLEKYGIEHPALIKK
Sbjct: 241 RLYDQLHVTFDSYNGEAFYNDKMDEVLDLLEAKNLLVESKGAQVVNLEKYGIEHPALIKK 300

Query: 312 SDGATLYITRDLAAALYRKRTYDFAKSIYVVGNEQSAHFKQLKAVLKEMDYDWSDDMTHV 371
           SDGATLYITRDLAAALYRKRTYDFAKS+YVVGNEQ+AHFKQLKAVLKEM YDWSDDMTHV
Sbjct: 301 SDGATLYITRDLAAALYRKRTYDFAKSVYVVGNEQAAHFKQLKAVLKEMGYDWSDDMTHV 360

Query: 372 PFGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADKDKVAQAVGVGAIK 431
            FGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADK+ VA AVGVGAIK
Sbjct: 361 AFGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADKEAVAHAVGVGAIK 420

Query: 432 FYDLKTDRTNGYDFDLEAMVSFEGETGPYVQYAHARIQSILRKANFSPSNSDNYSLNDVE 491
           FYDLKTDR NGYDFDLEAMVSFEGETGPYVQYAHARIQSILRKA+F+PS +  YSL D E
Sbjct: 421 FYDLKTDRMNGYDFDLEAMVSFEGETGPYVQYAHARIQSILRKADFTPSATTTYSLADAE 480

Query: 492 SWEIIKLIQDFPRIIVRAADNFEPSIIAKFAINLAQCFNKYYAHTRILDEDAEISSRLAL 551
           SWEIIKLIQDFPRII R +DNFEPSI+AKFAINLAQ FNKYYAHTRILD+++E  +RLAL
Sbjct: 481 SWEIIKLIQDFPRIIKRTSDNFEPSIMAKFAINLAQSFNKYYAHTRILDDNSERDNRLAL 540

Query: 552 CYATATVLKESLRLLGVDAPNEM 574
           CYATATVLKE+LRLLGVDAPNEM
Sbjct: 541 CYATATVLKEALRLLGVDAPNEM 563
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1076

A DNA sequence (GBSx1150) was identified in *S.agalactiae* <SEQ ID 3315> which encodes the amino acid sequence <SEQ ID 3316>. This protein is predicted to be arginine hydroximate resistance protein (argR). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3252(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10269> which encodes amino acid sequence <SEQ ID 10270> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA88596 GB:M18729 unknown protein [Streptococcus pneumoniae]
 Identities = 63/141 (44%), Positives = 90/141 (63%)

Query: 4   MNKIERQKRIKRLIQSGQIGTQEEIKLHLKNEGIDVTQATLSRDLREIGLLKLRSPEGKL 63
           M K +R + IK++I   ++ TQ+EI+  L+   + VTQ TLSRDLREIGL K++  +
Sbjct: 1   MRKRDRHQLIKKMITEEKLSTQKEIQDRLEAHNVCVTQTTLSRDLREIGLTKVKKNDMVY 60

Query: 64  YYSLSTATSNRFSPALRSYILKVSRASFMLVLNTNLGEASVLANFIDEKGLPEILGTMAG 123
           Y  ++          L  ++  V+RA F LVL+T LGEASVLAN +D      ILGT+AG
Sbjct: 61  YVLVNETEKIDLVEFLSHHLEGVARAEFTLVLHTKLGEASVLANIVDVNKDEWILGTVAG 120

Query: 124 ADTLLVICQNEDIAKVFEKEL 144
           A+TLLVIC+++ +AK+ E  L
Sbjct: 121 ANTLLVICRDQHVAKLMEDRL 141
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3317> which encodes the amino acid sequence <SEQ ID 3318>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3176(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 101/145 (69%), Positives = 121/145 (82%)

Query: 4   MNKIERQKRIKRLIQSGQIGTQEEIKLHLKNEGIDVTQATLSRDLREIGLLKLRSPEGKL 63
           MNK+ERQ++IKR+IQ+  IGTQE+IK HL+ EGI VTQATLSRDLREIGLLKLR  +GKL
Sbjct: 1   MNKMERQQQIKRIIQAEHIGTQEDIKNHLQKEGIVVTQATLSRDLREIGLLKLRDEQGKL 60

Query: 64  YYSLSTATSNRFSPALRSYILKVSRASFMLVLNTNLGEASVLANFIDEKGLPEILGTMAG 123
           YYSLS   +  FSP +R Y+LKV RA FMLVL+TNLGEA VLAN ID   + +ILGT+AG
Sbjct: 61  YYSLSEPVATPFSPEVRFYVLKVDRAGFMLVLHTNLGEADVLANLIDNDAIEDILGTIAG 120

Query: 124 ADTLLVICQNEDIAKVFEKELSVGL 148
           ADTLLVIC++E+IAK FEK+L+ GL
Sbjct: 121 ADTLLVICRDEEIAKRFEKDLAAGL 145
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1077

A DNA sequence (GBSx1151) was identified in *S.agalactiae* <SEQ ID 3319> which encodes the amino acid sequence <SEQ ID 3320>. This protein is predicted to be DNA mismatch repair protein hexa (mutS). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3570(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA88597 GB:M18729 mismatch repair protein [Streptococcus pneumoniae]
 Identities = 593/858 (69%), Positives = 698/858 (81%), Gaps = 14/858 (1%)

Query: 1   MAKPTISPGMQQYLDIKENYPDAFLLFRMGDFYELFYDDAVKAAQILEISLTSRNKNAEK 60
           MA   +SPGMQQY+DIK+ YPDAFLLFRMGDFYELFY+DAV AAQILEISLTSRNKNA+
Sbjct: 1   MAIEKLSPGMQQYVDIKKQYPDAFLLFRMGDFYELFYEDAVNAAQILEISLTSRNKNADN 60

Query: 61  PIPMAGVPYHSAQQYIDVLVELGYKVAIAEQMEDPKKAVGVVKREVVQVVTPGTVVESTK 120
           PIPMAGVPYHSAQQYIDVL+E GYKVAIAEQMEDPK+AVGVVKREVVQV+TPGTVV+S+K
Sbjct: 61  PIPMAGVPYHSAQQYIDVLIEQGYKVAIAEQMEDPKQAVGVVKREVVQVITPGTVVDSSK 120

Query: 121 PDSANNFLVAIDSQDQQTFGLAYMDVSTGEFQATLLTDFESVRSEILNLKAREIVVGYQL 180
           PDS NNFLV+ID +  Q FGLAYMD+ TG+F  T L DF  V  EI NLKARE+V+GY L
Sbjct: 121 PDSQNNFLVSIDREGNQ-FGLAYMDLVTGDFYVTGLLDFTLVCGEIRNLKAREVVLGYDL 179

Query: 181 TDEKNHLLTKQMNLLLSYEDERLNDIHLIDEQLTDLEISAAEKLLQYVHRTQKRELSHLQ 240
           ++E+  +L++QMNL+LSYE E   D+HL+D +L  +E +A+ KLLQYVHRTQ REL+HL+
```

```
Sbjct: 180 SEEEEQILSRQMNLVLSYEKESFEDLHLLDLRLATVEQTASSKLLQYVHRTQMRELNHLK 239

Query: 241 KVVHYEIKDYLQMSYATKNSLDLLENARTSKKHGSLYWLLDETKTAMGTRMLRTWIDRPL 300
            V+ YEIKD+LQM YATK SLDL+ENAR+ KK GSL+WLLDETKTAMG R+LR+WI RPL
Sbjct: 240 PVIRYEIKDFLQMDYATKASLDLVENARSGKKQGSLFWLLDETKTAMGMRLLRSWIHRPL 299

Query: 301 VSMNRIKERQDIIQVFLDYFFERNDLTESLKGVYDIERLASRVSFGKANPKDLLQLGQTL 360
           +   RI +RQ+++QVFLD+FFER+DLT+SLKGVYDIERLASRVSFGK NPKDLLQL  TL
Sbjct: 300 IDKERIVQRQEVVQVFLDHFFERSDLTDSLKGVYDIERLASRVSFGKTNPKDLLQLATTL 359

Query: 361 SQIPRIKMILQSFNQPELDIIVNKIDTMPELESLINTAIAPEAQATITEGNIIKSGFDKQ 420
           S +PRI+ IL+   QP L  ++ ++D +PELESLI+ AIAPEA   IT+G II++GFD+
Sbjct: 360 SSVPRIRAILEGMEQPTLAYLIAQLDAIPELESLISAAIAPEAPHVITDGGIIRTGFDET 419

Query: 421 LDNYRTVMREGTGWIADIEAKERAASGIGTLKIDYNKKDGYYFHVTNSNLSLVPEHFFRK 480
           LD YR V+REGT WIA+IEAKER  SGI TLKIDYNKKDGYYFHVTNS L  VP HFFRK
Sbjct: 420 LDKYRCVLREGTSWIAEIEAKERENSGISTLKIDYNKKDGYYFHVTNSQLGNVPAHFFRK 479

Query: 481 ATLKNSERYGTAELAKIEGEMLEAREQSSNLEYDIFMRVRAQVESYIKRLQELAKTIATV 540
           ATLKNSER+GT ELA+IEG+MLEARE+S+NLEY+IFMR+R +V  YI+RLQ LA+ IATV
Sbjct: 480 ATLKNSERFGTEELARIEGDMLEAREKSANLEYEIFMRIREEVGKYIQRLQALAQGIATV 539

Query: 541 DVLQSLAVVAENYHYVRPKFNDQHQIKIKNGRHATVEKVMGVQEYIPNSIYFDSQTDIQL 600
           DVLQSLAVVAE  H +RP+F D  QI I+ GRHA VEKVMG Q YIPN+I      T IQL
Sbjct: 540 DVLQSLAVVAETQHLIRPEFGDDSQIDIRKGRHAVVEKVMGAQTYIPNTIQMAEDTSIQL 599

Query: 601 ITGPNMSGKSTYMRQLALTVIMAQMGGFVSADEVDLPVFDAIFTRIGAADDLISGQSTFM 660
           +TGPNMSGKSTYMRQLA+T +MAQ+G +V A+   LP+FDAIFTRIGAADDL+SGQSTFM
Sbjct: 600 VTGPNMSGKSTYMRQLAMTAVMAQLGSYVPAESAHLPIFDAIFTRIGAADDLVSGQSTFM 659

Query: 661 VEMMEANQAVKRASDKSLILFDELGRGTATYDGMALAQSIIEYIHDRVRAKTMFATHYHE 720
           VEMMEAN A+  A+  SLILFDELGRGTATYDGMALAQSIIEYIH+ + AKT+FATHYHE
Sbjct: 660 VEMMEANNAISHATKNSLILFDELGRGTATYDGMALAQSIIEYIHEHIGAKTLFATHYHE 719

Query: 721 LTDLSEQLTRLVNVHVATLERDGEVTFLHKIESGPADKSYGIHVAKIAGLPIDLLDRATD 780
           LT L   L  LVNVHVATLE+DG+VTFLHKIE GPADKSYGIHVAKIAGLP DLL RA
Sbjct: 720 LTSLESSLQHLVNVHVATLEQDGQVTFLHKIEPGPADKSYGIHVAKIAGLPADLLARADK 779

Query: 781 ILSQLEADAVQLIVSPSQEAVTADLNEELDSEKQQGQLSLFEEPSNAGRVIEELEAIDIM 840
           IL+QLE    +   SP     T+ + E         Q+SLF+  +    ++ EL  +D+
Sbjct: 780 ILTQLENQGTE---SPPPMRQTSAVTE---------QISLFDR-AEEHPILAELAKLDVY 826

Query: 841 NLTPMQAMNAIFDLKKLL 858
           N+TPMQ MN + +LK+ L
Sbjct: 827 NMTPMQVMNVLVELKQKL 844
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3321> which encodes the amino acid sequence <SEQ ID 3322>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38    Transmembrane  532 - 548 ( 532 - 549)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 661/858 (77%), Positives = 746/858 (86%), Gaps = 7/858 (0%)

Query: 1   MAKPTISPGMQQYLDIKENYPDAFLLFRMGDFYELFYDDAVKAAQILEISLTSRNKNAEK 60
           MAK  ISPGMQQYLDIK++YPDAFLLFRMGDFYELFY+DAVKAAQ+LEI LTSRNKNAE
Sbjct: 1   MAKTNISPGMQQYLDIKKDYPDAFLLFRMGDFYELFYEDAVKAAQLLEIGLTSRNKNAEN 60

Query: 61  PIPMAGVPYHSAQQYIDVLVELGYKVAIAEQMEDPKKAVGVVKREVVQVVTPGTVVESTK 120
```

```
           PIPMAGVP+HSAQQYIDVL+ELGYKVA+AEQMEDPK+AVGVVKREVVQV+TPGTVV+S K
Sbjct: 61  PIPMAGVPHHSAQQYIDVLIELGYKVAVAEQMEDPKQAVGVVKREVVQVITPGTVVDSAK 120

Query: 121 PDSANNFLVAIDSQDQQTFGLAYMDVSTGEFQATLLTDFESVRSEILNLKAREIVVGYQL 180
           PDSANNFLVA+D  D   +GLAYMDVSTGEF  T L DF SVRSEI NLKA+E+++G+ L
Sbjct: 121 PDSANNFLVAVDF-DGCRYGLAYMDVSTGEFCVTDLADFTSVRSEIQNLKAKEVLLGFDL 179

Query: 181 TDEKNHLLTKQMNLLLSYEDERLNDIHLIDEQLTDLEISAAEKLLQYVHRTQKRELSHLQ 240
           ++E+  +L KQMNLLLSYE+    D  LID QLT +E++AA KLLQYVH+TQ RELSHLQ
Sbjct: 180 SEEEQTILVKQMNLLLSYEETVYEDKSLIDGQLTTVELTAAGKLLQYVHKTQMRELSHLQ 239

Query: 241 KVVHYEIKDYLQMSYATKNSLDLLENARTSKKHGSLYWLLDETKTAMGTRMLRTWIDRPL 300
            +VHYEIKDYLQMSYATK+SLDL+ENART+KKHGSLYWLLDETKTAMG R+LR+WIDRPL
Sbjct: 240 ALVHYEIKDYLQMSYATKSSLDLVENARTNKKHGSLYWLLDETKTAMGMRLLRSWIDRPL 299

Query: 301 VSMNRIKERQDIIQVFLDYFFERNDLTESLKGVYDIERLASRVSFGKANPKDLLQLGQTL 360
           VS   I ERQ+IIQVFL+ F ER DL+ SLKGVYDIERL+SRVSFGKANPKDLLQLG TL
Sbjct: 300 VSKEAILERQEIIQVFLNAFIERTDLSNSLKGVYDIERLSSRVSFGKANPKDLLQLGHTL 359

Query: 361 SQIPRIKMILQSFNQPELDIIVNKIDTMPELESLINTAIAPEAQATITEGNIIKSGFDKQ 420
           +Q+P IK IL+SF+ P +D +VN ID++PELE LI TAI P+A ATI+EG+II++GFD++
Sbjct: 360 AQVPYIKAILESFDSPCVDKLVNDIDSLPELEYLIRTAIDPDAPATISEGSIIRNGFDER 419

Query: 421 LDNYRTVMREGTGWIADIEAKERAASGIGTLKIDYNKKDGYYFHVTNSNLSLVPEHFFRK 480
           LD+YR VMREGTGWIADIEAKER ASGI  LKIDYNKKDGYYFHVTNSNLSLVPEHFFRK
Sbjct: 420 LDHYRKVMREGTGWIADIEAKERQASGINNLKIDYNKKDGYYFHVTNSNLSLVPEHFFRK 479

Query: 481 ATLKNSERYGTAELAKIEGEMLEAREQSSNLEYDIFMRVRAQVESYIKRLQELAKTIATV 540
           ATLKNSERYGTAELAKIEG+MLEARE+SS+LEYDIFM +RAQVE+YI RLQ+LAK +ATV
Sbjct: 480 ATLKNSERYGTAELAKIEGQMLEAREESSSLEYDIFMCIRAQVETYINRLQKLAKILATV 539

Query: 541 DVLQSLAVVAENYHYVRPKFNDQHQIKIKNGRHATVEKVMGVQEYIPNSIYFDSQTDIQL 600
           DVLQSLAVVAE  HY+RP+FND H I I+ GRHA VEKVMGVQEYIPNSI FD QT IQL
Sbjct: 540 DVLQSLAVVAETNHYIRPQFNDNHVITIQEGRHAVVEKVMGVQEYIPNSISFDQQTSIQL 599

Query: 601 ITGPNMSGKSTYMRQLALTVIMAQMGGFVSADEVDLPVFDAIFTRIGAADDLISGQSTFM 660
           ITGPNMSGKSTYMRQLALTVIMAQMG FV+AD VDLP+FDAIFTRIGAADDLISGQSTFM
Sbjct: 600 ITGPNMSGKSTYMRQLALTVIMAQMGSFVAADHVDLPLFDAIFTRIGAADDLISGQSTFM 659

Query: 661 VEMMEANQAVKRASDKSLILFDELGRGTATYDGMALAQSIIEYIHDRVRAKTMFATHYHE 720
           VEMMEANQA+KRASD SLILFDELGRGTATYDGMALAQ+IIEYIHDRV AKT+FATHYHE
Sbjct: 660 VEMMEANQAIKRASDNSLILFDELGRGTATYDGMALAQAIIEYIHDRVGAKTIFATHYHE 719

Query: 721 LTDLSEQLTRLVNVHVATLERDGEVTFLHKIESGPADKSYGIHVAKIAGLPIDLLDRATD 780
           LTDLS  LT LVNVHVATLE+DG+VTFLHKI  GPADKSYGIHVAKIAGLP  LL RA +
Sbjct: 720 LTDLSTNLTSLVNVHVATLEKDGDVTFLHKIAEGPADKSYGIHVAKIAGLPKSLLKRADE 779

Query: 781 ILSQLEADAVQLIVSPSQEAVTADLNEELDSEKQQGQLSLFEEPSNAGRVIEELEAIDIM 840
           +L++LE        S S E ++     E  S  +QGQLSLF +   A  + + LE ID+M
Sbjct: 780 VLTRLETQ------SRSTEIISVPSQVESSSAVRQGQLSLFGDEEKAHEIRQALEVIDVM 833

Query: 841 NLTPMQAMNAIFDLKKLL 858
           N+TP+QAM  +++LKKLL
Sbjct: 834 NMTPLQAMTTLYELKKLL 851
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1078

A DNA sequence (GBSx1152) was identified in *S.agalactiae* <SEQ ID 3323> which encodes the amino acid sequence <SEQ ID 3324>. This protein is predicted to be cold shock protein-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2095(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB69404 GB:A91080 unnamed protein product [unidentified]
 Identities = 48/63 (76%), Positives = 56/63 (88%)

Query: 1  MTQGTVKWFNSEKGFGFISSETGTDVFAHFSEIKVDGFKTLEEGQKVTFDIQDGQRGPQA 60
          MT+GTVKWFN +KGFGFI+SE G DVFAHFS+I+  GFKTL+EGQKVTFD++ GQRGPQA
Sbjct: 1  MTKGTVKWFNPDKGFGFITSEDGQDVFAHFSQIQTSGFKTLDEGQKVTFDVEAGQRGPQA 60

Query: 61 TNI 63
           NI
Sbjct: 61 VNI 63
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3325> which encodes the amino acid sequence <SEQ ID 3326>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2350(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 49/63 (77%), Positives = 56/63 (88%)

Query: 1  MTQGTVKWFNSEKGFGFISSETGTDVFAHFSEIKVDGFKTLEEGQKVTFDIQDGQRGPQA 60
          M QGTVKWFN+EKGFGFIS+E G DVFAHFS I+ +GFKTLEEGQKV FD+++GQRGPQA
Sbjct: 3  MAQGTVKWFNAEKGFGFISTENGQDVFAHFSAIQTNGFKTLEEGQKVAFDVEEGQRGPQA 62

Query: 61 TNI 63
           NI
Sbjct: 63 VNI 65
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1079

A DNA sequence (GBSx1153) was identified in *S.agalactiae* <SEQ ID 3327> which encodes the amino acid sequence <SEQ ID 3328>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6378(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1080

A DNA sequence (GBSx1154) was identified in *S.agalactiae* <SEQ ID 3329> which encodes the amino acid sequence <SEQ ID 3330>. This protein is predicted to be DNA mismatch repair protein hexb (mutL). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2242(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10267> which encodes amino acid sequence <SEQ ID 10268> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA88600 GB:M29686 mismatch repair protein [Streptococcus pneumoniae]
 Identities = 452/657 (68%), Positives = 543/657 (81%), Gaps = 8/657 (1%)

Query: 20  LSKIIELPDILANQIAAGEVVERPSSVVKELVENAIDAGSSQITIEVEESGLKKIQITDN 79
            +S IIELP++LANQIAAGEV+ERP+SV KELVENAIDAGSSQI IE+EE+GLKK+QITDN
Sbjct: 1   MSHIIELPEMLANQIAAGEVIERPASVCKELVENAIDAGSSQIIIEIEEAGLKKVQITDN 60

Query: 80  GEGMTSEDAVLSLRRHATSKIKSQSDLFRIRTLGFRGEALPSIASISLMTIKTATEQGKQ 139
            G G+  ++  L+LRRHATSKIK+Q+DLFRIRTLGFRGEALPSIAS+S++T+ TA +
Sbjct: 61  GHGIAHDEVELALRRHATSKIKNQADLFRIRTLGFRGEALPSIASVSVLTLLTAVDGASH 120

Query: 140 GTLLVAKGGNIEKQEVVSSPRGTKILVENLFFNTPARLKYMKSLQSELAHIIDIVNRLSL 199
            GT LVA+GG +E+    +SP GTK+ VE+LFFNTPARLKYMKS Q+EL+HIIDIVNRL L
Sbjct: 121 GTKLVARGGEVEEVIPATSPVGTKVCVEDLFFNTPARLKYMKSQQAELSHIIDIVNRLGL 180

Query: 200 AHPEVAFTLINDGKEMTKTSGTGDLRQAIAGIYGLNTAKKMIEISNADLDFEISGYVSLP 259
            AHPE++F+LI+DGKEMT+T+GTG LRQAIAGIYGL +AKKMIEI N+DLDFEISG+VSLP
Sbjct: 181 AHPEISFSLISDGKEMTRTAGTGQLRQAIAGIYGLVSAKKMIEIENSDLDFEISGFVSLP 240

Query: 260 ELTRANRNYITLLINGRYIKNFLLNRSILDGYGSKLMVGRFPIAVIDIQIDPYLADVNVH 319
            ELTRANRNYI+L INGRYIKNFLLNR+ILDG+GSKLMVGRFP+AVI I IDPYLADVNVH
Sbjct: 241 ELTRANRNYISLFINGRYIKNFLLNRAILDGFGSKLMVGRFPLAVIHIHIDPYLADVNVH 300

Query: 320 PTKQEVRISKERELMSLISTAISESLKQYDLIPDALENLAKTSTRSVDKPIQTSFSLKQP 379
            PTKQEVRISKE+ELM+L+S AI+ SLK+  LIPDALENLAK++ R+ +K  QT   LK+
Sbjct: 301 PTKQEVRISKEKELMTLVSEAIANSLKEQTLIPDALENLAKSTVRNREKVEQTILPLKEN 360

Query: 380 GLYYDRAKNDFFIGADTVSEPIANFTNLDKSDGSVDNDVKNSVNQGATQSPNIKYASRDQ 439
             LYY++ +        + +E       L      +    K ++++  T+   + +A R
Sbjct: 361 TLYYEKTEP----SRPSQTEVADYQVELTDEGQDLTLFAKETLDR-LTKPAKLHFAERKP 415

Query: 440 ADSENFIHSQDYLSSKQSLNKLVEKLDSEESSTFPELEFFGQMHGTYLFAQGNGGLYIID 499
            A+ +   H +  L+   S++K  +KL+ EE+S+FPELEFFGQMHGTYLFAQG  GLYIID
Sbjct: 416 ANYDQLDHPELDLA---SIDKAYDKLEREEASSFPELEFFGQMHGTYLFAQGRDGLYIID 472

Query: 500 QHAAQERVKYEYYREKIGEVDNSLQQLLVPFLFEFSSSDFLQLQEKMSLLQDVGIFLEPY 559
            QHAAQERVKYE YRE IG VD S QQLLVP++FEF + D L+L+E+M LL++VG+FL  Y
Sbjct: 473 QHAAQERVKYEEYRESIGNVDQSQQQLLVPYIFEFPADDALRLKERMPLLEEVGVFLAEY 532

Query: 560 GNNTFILREHPIWMKEEEVESGIYEMCDMLLLTNEVSVKKYRAELAIMMSCKRSIKANHT 619
            G N FILREHPIWM EEE+ESGIYEMCDMLLLT EVS+KKYRAELAIMMSCKRSIKANH
Sbjct: 533 GENQFILREHPIWMAEEEIESGIYEMCDMLLLTKEVSIKKYRAELAIMMSCKRSIKANHR 592
```

```
Query: 620 LDDYSARHLLDQLAQCKNPYNCPHGRPVLVNFTKADMEKMFKRIQENHTSLRDLGKY 676
           +DD+SAR LL QL+QC NPYNCPHGRPVLV+FTK+DMEKMF+RIQENHTSLR+LGKY
Sbjct: 593 IDDHSARQLLYQLSQCDNPYNCPHGRPVLVHFTKSDMEKMFRRIQENHTSLRELGKY 649
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3331> which encodes the amino acid sequence <SEQ ID 3332>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1854(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 502/663 (75%), Positives = 574/663 (85%), Gaps = 9/663 (1%)

Query: 20  LSKIIELPDILANQIAAGEVVERPSSVVKELVENAIDAGSSQITIEVEESGLKKIQITDN 79
           ++ IIELP++LANQIAAGEVVERP+SVVKELVENAIDA SSQIT+E+EESGLK IQ+TDN
Sbjct: 14  MTNIIELPEVLANQIAAGEVVERPASVVKELVENAIDAKSSQITVEIEESGLKMIQVTDN 73

Query: 80  GEGMTSEDAVLSLRRHATSKIKSQSDLFRIRTLGFRGEALPSIASISLMTIKTATEQGKQ 139
           GEGM+ ED  LSLRRHATSKIKSQSDLFRIRTLGFRGEALPS+ASIS +TIKTAT++
Sbjct: 74  GEGMSHEDLPLSLRRHATSKIKSQSDLFRIRTLGFRGEALPSVASISKITIKTATKEVTH 133

Query: 140 GTLLVAKGGNIEKQEVVSSPRGTKILVENLFFNTPARLKYMKSLQSELAHIIDIVNRLSL 199
           G+LL+A GG IE  E +S+P GTKI VENLF+NTPARLKYMKSLQ+ELAHI+D+VNRLSL
Sbjct: 134 GSLLIATGGEIETLEAISTPTGTKIKVENLFYNTPARLKYMKSLQAELAHIVDVVNRLSL 193

Query: 200 AHPEVAFTLINDGKEMTKTSGTGDLRQAIAGIYGLNTAKKMIEISNADLDFEISGYVSLP 259
           AHPEVAFTLI+DG+++T+TSGTGDLRQAIAGIYGLNT KKM+ ISNADLDFE+SGYVSLP
Sbjct: 194 AHPEVAFTLISDGRQLTQTSGTGDLRQAIAGIYGLNTTKKMLAISNADLDFEVSGYVSLP 253

Query: 260 ELTRANRNYITLLINGRYIKNFLLNRSILDGYGSKLMVGRFPIAVIDIQIDPYLADVNVH 319
           ELTRANRNY+T+L+NGRYIKNFLLNR+ILDGYGSKLMVGRFPI VIDIQIDPYLADVNVH
Sbjct: 254 ELTRANRNYMTILVNGRYIKNFLLNRAILDGYGSKLMVGRFPIVVIDIQIDPYLADVNVH 313

Query: 320 PTKQEVRISKERELMSLISTAISESLKQYDLIPDALENLAKTSTRSVDKPIQTSFSLKQP 379
           PTKQEVRISKERELM+LISTAISESLK+ DLIPDALENLAK+STR   KP QT   L+
Sbjct: 314 PTKQEVRISKERELMALISTAISESLKEQDLIPDALENLAKSSTRHFSKPEQTQLPLQSR 373

Query: 380 GLYYDRAKNDFFIGADTVSEPIANFTNLDKSDGSVDNDVKNSV------NQGATQSPNIK 433
           GLYYD  KNDFF+    VSE I      D   G+VDN VK         ++      ++K
Sbjct: 374 GLYYDPQKNDFFVKESAVSEKI---PETDFYSGAVDNSVKVEKVELLPHSEEVIGPSSVK 430

Query: 434 YASRDQADSENFIHSQDYLSSKQSLNKLVEKLDSEESSTFPELEFFGQMHGTYLFAQGNG 493
           +ASR Q       H    L ++Q L++++ +L++E  S FPEL++FGQMHGTYLFAQG
Sbjct: 431 HASRPQNTFTETDHPNLDLKNRQKLSQMLTRLENEGQSVFPELDYFGQMHGTYLFAQGKD 490

Query: 494 GLYIIDQHAAQERVKYEYYREKIGEVDNSLQQLLVPFLFEFSSSDFLQLQEKMSLLQDVG 553
           GL+IIDQHAAQERVKYEYYR+KIGEVD+SLQQLLVP+LFEFS SDF+ LQEKM+LL +VG
Sbjct: 491 GLFIIDQHAAQERVKYEYYRDKIGEVDSSLQQLLVPYLFEFSGSDFINLQEKMALLNEVG 550

Query: 554 IFLEPYGNNTFILREHPIWMKEEEVESGIYEMCDMLLLTNEVSVKKYRAELAIMMSCKRS 613
           IFLE YG+NTFILREHPIWMKEEE+ SG+YEMCDMLLLTNEVS+K YRAELAIMMSCKRS
Sbjct: 551 IFLEVYGHNTFILREHPIWMKEEEIASGVYEMCDMLLLTNEVSIKTYRAELAIMMSCKRS 610

Query: 614 IKANHTLDDYSARHLLDQLAQCKNPYNCPHGRPVLVNFTKADMEKMFKRIQENHTSLRDLGKY 676
           IKANH+LDDYSAR+LL QLAQC+NPYNCPHGRPVL+NF+KADMEKMF+RIQENHTSLR+LGKY
Sbjct: 611 IKANHSLDDYSARNLLLQLAQCQNPYNCPHGRPVLINFSKADMEKMFRRIQENHTSLRELGKY 673
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1081

A DNA sequence (GBSx1155) was identified in *S.agalactiae* <SEQ ID 3333> which encodes the amino acid sequence <SEQ ID 3334>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3372(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1082

A DNA sequence (GBSx1156) was identified in *S.agalactiae* <SEQ ID 3335> which encodes the amino acid sequence <SEQ ID 3336>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.01    Transmembrane  176 - 192 ( 170 - 197)
   INTEGRAL    Likelihood = -8.07    Transmembrane  390 - 406 ( 387 - 412)
   INTEGRAL    Likelihood = -6.10    Transmembrane  271 - 287 ( 269 - 291)
   INTEGRAL    Likelihood = -6.00    Transmembrane   83 -  99 (  82 - 101)
   INTEGRAL    Likelihood = -4.78    Transmembrane   51 -  67 (  50 -  71)
   INTEGRAL    Likelihood = -2.92    Transmembrane  303 - 319 ( 302 - 320)
   INTEGRAL    Likelihood = -2.76    Transmembrane  363 - 379 ( 362 - 381)
   INTEGRAL    Likelihood = -2.39    Transmembrane  152 - 168 ( 151 - 169)
   INTEGRAL    Likelihood = -2.02    Transmembrane  325 - 341 ( 325 - 342)
   INTEGRAL    Likelihood = -1.65    Transmembrane  226 - 242 ( 226 - 242)
   INTEGRAL    Likelihood = -0.90    Transmembrane   24 -  40 (  24 -  40)
   INTEGRAL    Likelihood = -0.27    Transmembrane  111 - 127 ( 111 - 127)

----- Final Results -----
               bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10265> which encodes amino acid sequence <SEQ ID 10266> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA61918 GB:X89779 LmrP integral membrane protein [Lactococcus
           lactis]
 Identities = 145/401 (36%), Positives = 236/401 (58%), Gaps = 4/401 (0%)

Query: 9   VKEFFALPKQLQLRELLRFISITVGSAIFPFMAMYYVQYFGNLVTGILIIITQLSGFVAT 68
           +KEF+ L K LQLR  + F+       +F  M +YY QY G+ +TGIL+ ++ ++ FVA
Sbjct: 1   MKEFWNLDKNLQLRLGIVFLGAFSYGTVFSSMTIYYNQYLGSAITGILLALSAVATFVAG 60

Query: 69  LYGGHLSDAMGRKKVVIIGSLLATIGWAITIAANVPNHITPHLTFVGILIIEIAHQFYFP 128
           +  G  +D  GRK V++ G+++  +G A+ IA+N+P H+ P  TF+  L+I    + F
Sbjct: 61  ILAGFFADRNGRKPVMVFGTIIQLLGAALAIASNLPGHVNPWSTFIAFLLISFGYNFVIT 120

Query: 129 AYEAMTIDLTNEQNRRFVYTIGYWLVNIAVMLGSGIAGIFYDHHFFELLIVLLIISAICC 188
           A  AM ID +N +NR+ V+ + YW  N++V+LG+ +    +   F  LL++LL+   +
```

```
Sbjct: 121 AGNAMIIDASNAENRKVVFMLDYWAQNLSVILGAALGAWLFRPAFEALLVILLLTVLVSF 180

Query: 189 FVVYFKFDET-KPQEGTFKHDKGVLGTFKNYSQVLVDKAFVVYTLGAIGSSVVWLQVDNY 247
           F+  F   ET KP   T K D+     F+ Y  VL DK ++++    I ++ + +Q DN+
Sbjct: 181 FLTTFVMTETFKP---TVKVDEKAENIFQAYKTVLQDKTYMIFMGANIATTFIIMQFDNF 237

Query: 248 FSVNLKQNFEVVSILGHTITGAKMLSLAVFTNTLLIVLLMTTINKFIENWPLKRQLILGS 307
             V+L  +F+ ++  G  I G +ML++ +    +L+VLLMTT+N+  ++W  ++  I GS
Sbjct: 238 LPVHLSNSFKTITFWGFEIYGQRMLTIYLILACVLVVLLMTTLNRLTKDWSHQKGFIWGS 297

Query: 308 LICGFGMLFNISLNTFGAILIAMTFFTFGEMIYVPASQVLRAEMMVEGKIGSYSGFLAIA 367
           L    GM+F+    TF  I IA   +T GE++Y P+ Q L A++M   KIGSY+G  AI
Sbjct: 298 LFMAIGMIFSFLTTTFTPIFIAGIVYTLGEIVYTPSVQTLGADLMNPEKIGSYNGVAAIK 357

Query: 368 QPVASVLAGAMVSLSYFTGKIGVQITLTIFMLAGLVLILYA 408
            P+AS+LAG +VS+S     IGV + L +  +  ++L+L A
Sbjct: 358 MPIASILAGLLVSISPMIKAIGVSLVLALTEVLAIILVLVA 398
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3337> which encodes the amino acid sequence <SEQ ID 3338>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.41   Transmembrane  166 - 182 ( 161 - 188)
   INTEGRAL    Likelihood = -7.75   Transmembrane  384 - 400 ( 376 - 403)
   INTEGRAL    Likelihood = -7.64   Transmembrane  266 - 282 ( 261 - 285)
   INTEGRAL    Likelihood = -4.25   Transmembrane  295 - 311 ( 291 - 313)
   INTEGRAL    Likelihood = -2.71   Transmembrane   98 - 114 (  98 - 115)
   INTEGRAL    Likelihood = -2.23   Transmembrane  355 - 371 ( 355 - 374)
   INTEGRAL    Likelihood = -2.02   Transmembrane  218 - 234 ( 218 - 234)
   INTEGRAL    Likelihood = -1.91   Transmembrane  315 - 331 ( 315 - 331)
   INTEGRAL    Likelihood = -1.22   Transmembrane   75 -  91 (  75 -  92)
   INTEGRAL    Likelihood = -0.75   Transmembrane   45 -  61 (  45 -  63)
   INTEGRAL    Likelihood = -0.75   Transmembrane  144 - 160 ( 144 - 161)

----- Final Results -----
                bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA61918 GB:X89779 LmrP integral membrane protein [Lactococcus
           lactis]
 Identities = 138/400 (34%), Positives = 223/400 (55%), Gaps = 2/400 (0%)

Query: 1   MQEFLNLPKQIQLRQLVRFVTITLGSSIFPFMAMYYTTYFGTFWTGLLMMITSLMGFVGT 60
           M+EF NL K +QLR  + F+      ++F  M +YY  Y G+  TG+L+ ++++  FV
Sbjct: 1   MKEFWNLDKNLQLRLGIVFLGAFSYGTVFSSMTIYYNQYLGSAITGILLALSAVATFVAG 60

Query: 61  LYGGHLSDALGRKKVIMIGSVGTTLGWFLTILANLPNAAIPWLTFAGILLVEIASSFYGP 120
           +  G  +D  GRK V++ G++   LG  L I +NLP    PW TF   LL+    +F
Sbjct: 61  ILAGFFADRNGRKPVMVFGTIIQLLGAALAIASNLPGHVNPWSTFIAFLLISFGYNFVIT 120

Query: 121 AYEAMLIDLTDESNRRFVYTINYWFINIAVMFGAGLSGLFYDHHFLALLVALLLVNVLCF 180
           A  AM+ID ++  NR+ V+ ++YW  N++V+ GA L   +    F ALLV LLL ++  F
Sbjct: 121 AGNAMIIDASNAENRKVVFMLDYWAQNLSVILGAALGAWLFRPAFEALLVILLLTVLVSF 180

Query: 181 GVAYYCFDETRPETHAFDHGKGLLASFQNYRQVFHDRAFVLFTLGAIFSGSIWMQMDNYV 240
            +  +   ET   T   D     +  FQ Y+ V  D+ +++F    I +  I MQ DN++
Sbjct: 181 FLTTFVMTETFKPTVKVDEKAENI--FQAYKTVLQDKTYMIFMGANIATTFIIMQFDNFL 238

Query: 241 PVHLKLYFQPTAVLGFQVTSSKMLSLMVLTNTLLIVLFMTVVNKLTEKWKLLPQLVVGSL 300
           PVHL   F+     GF++   +ML++ ++   +L+VL MT +N+LT+ W      + GSL
Sbjct: 239 PVHLSNSFKTITFWGFEIYGQRMLTIYLILACVLVVLLMTTLNRLTKDWSHQKGFIWGSL 298

Query: 301 LFTLGMLLSFTFTQFYAIWLSVVLLTFGEMINVSASQVLRADMMDHSQIGSYTGFVSMAQ 360
```

```
              +GM+ SF  T F  I+++ ++ T GE++   + Q L AD+M+  +IGSY G  ++
Sbjct: 299 FMAIGMIFSFLTTTFTPIFIAGIVYTLGEIVYTPSVQTLGADLMNPEKIGSYNGVAAIKM 358

Query: 361 PLGAILASLLVSVSHFTGPLGVQCLFAVIALLGIYFTVVS 400
           P+ +ILA LLVS+S     +GV  + A+  +L I   +V+
Sbjct: 359 PIASILAGLLVSISPMIKAIGVSLVLALTEVLAIILVLVA 398
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 228/406 (56%), Positives = 305/406 (74%)

Query: 9   VKEFFALPKQLQLRELLRFISITVGSAIFPFMAMYYVQYFGNLVTGILIIITQLSGFVAT 68
           ++EF  LPKQ+QLR+L+RF++IT+GS+IFPFMAMYY  YFG   TG+L++IT L GFV T
Sbjct: 1   MQEFLNLPKQIQLRQLVRFVTITLGSSIFPFMAMYYTTYFGTFWTGLLMMITSLMGFVGT 60

Query: 69  LYGGHLSDAMGRKKVVIIGSLLATIGWAITIAANVPNHITPHLTFVGILIIEIAHQFYFP 128
           LYGGHLSDA+GRKKV++IGS+  T+GW +TI AN+PN   P LTF GIL++EIA  FY P
Sbjct: 61  LYGGHLSDALGRKKVIMIGSVGTTLGWFLTILANLPNAAIPWLTFAGILLVEIASSFYGP 120

Query: 129 AYEAMTIDLTNEQNRRFVYTIGYWLVNIAVMLGSGIAGIFYDHHFFELLIVLLIISAICC 188
           AYEAM IDLT+E NRRFVYTI YW +NIAVM G+G++G+FYDHHF  LL+ LL+++ +C
Sbjct: 121 AYEAMLIDLTDESNRRFVYTINYWFINIAVMFGAGLSGLFYDHHFLALLVALLLVNVLCF 180

Query: 189 FVVYFKFDETKPQEGTFKHDKGVLGTFKNYSQVLVDKAFVVYTLGAIGSSVVWLQVDNYF 248
            V Y+ FDET+P+   F H KG+L +F+NY QV  D+AFV++TLGAI S  +W+Q+DNY
Sbjct: 181 GVAYYCFDETRPETHAFDHGKGLLASFQNYRQVFHDRAFVLFTLGAIFSGSIWMQMDNYV 240

Query: 249 SVNLKQNFEVVSILGHTITGAKMLSLAVFTNTLLIVLLMTTINKFIENWPLKRQLILGSL 308
            V+LK  F+  ++LG  +T +KMLSL V TNTLLIVL MT +NK  E W L  QL++GSL
Sbjct: 241 PVHLKLYFQPTAVLGFQVTSSKMLSLMVLTNTLLIVLFMTVVNKLTEKWKLLPQLVVGSL 300

Query: 309 ICGFGMLFNISLNTFGAILIAMTFFTFGEMIYVPASQVLRAEMMVEGKIGSYSGFLAIAQ 368
           +   GML + +   F AI +++   TFGEMI V ASQVLRA+MM   +IGSY+GF+++AQ
Sbjct: 301 LFTLGMLLSFTFTQFYAIWLSVVLLTFGEMINVSASQVLRADMMDHSQIGSYTGFVSMAQ 360

Query: 369 PVASVLAGAMVSLSYFTGKIGVQITLTIFMLAGLVLILYATKMKNI 414
           P+ ++LA  +VS+S+FTG +GVQ    +  L G+   + + KMK +
Sbjct: 361 PLGAILASLLVSVSHFTGPLGVQCLFAVIALLGIYFTVVSAKMKKV 406
```

A related GBS gene <SEQ ID 8725> and protein <SEQ ID 8726> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
SRCFLG: 0
McG: Length  of  UR:   4
     Peak Value of UR:   1.73
     Net Charge of CR: 1
McG: Discrim Score:     -4.26
GvH: Signal Score (-7.5): -2.48
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 12 value: -14.01 threshold:  0.0
   INTEGRAL    Likelihood =-14.01   Transmembrane  168 - 184 ( 162 - 189)
   INTEGRAL    Likelihood = -8.07   Transmembrane  382 - 398 ( 379 - 404)
   INTEGRAL    Likelihood = -6.10   Transmembrane  263 - 279 ( 261 - 283)
   INTEGRAL    Likelihood = -6.00   Transmembrane   75 -  91 (  74 -  93)
   INTEGRAL    Likelihood = -4.78   Transmembrane   43 -  59 (  42 -  63)
   INTEGRAL    Likelihood = -2.92   Transmembrane  295 - 311 ( 294 - 312)
   INTEGRAL    Likelihood = -2.76   Transmembrane  355 - 371 ( 354 - 373)
   INTEGRAL    Likelihood = -2.39   Transmembrane  144 - 160 ( 143 - 161)
   INTEGRAL    Likelihood = -2.02   Transmembrane  317 - 333 ( 317 - 334)
   INTEGRAL    Likelihood = -1.65   Transmembrane  218 - 234 ( 218 - 234)
   INTEGRAL    Likelihood = -0.90   Transmembrane   16 -  32 (  16 -  32)
   INTEGRAL    Likelihood = -0.27   Transmembrane  103 - 119 ( 103 - 119)
   PERIPHERAL  Likelihood =  9.44     239
 modified ALOM score:   3.30
```

```
icml HYPID: 7  CFP: 0.660

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01675(325 - 1530 of 1854)
EGAD|40187|42348(1 - 400 of 408) integral membrane protein (lmrP) {Lactococcus lactis}
GP|1052754|emb|CAA61918.1||X89779 LmrP integral membrane protein {Lactococcus lactis}
PIR|S58131|S58131 integral membrane protein LmrP - Lactococcus lactis
%Match = 21.7
%Identity = 36.2  %Similarity = 60.8
Matches = 145  Mismatches = 155  Conservative Sub.s = 99

243       273       303       333       363       393       423       453
LQKLIWRKCLNESKKIIQASGI*ENIDNYLLGKKGEKVKEFFALPKQLQLRELLRFISITVGSAIFPFMAMYYVQYFGNL
                                     :|||: | | ||||  : |:       :|  | :|| ||:|:
                                     MKEFWNLDKNLQLRLGIVFLGAFSYGTVFSSMTIYYNQYLGSA
                                             10        20        30        40

483       513       543       573       603       633       663       693
VTGILIIITQLSGFVATLYGGHLSDAMGRKKVVIIGSLLATIGWAITIAANVPNHITPHLTFVGILIIEIAHQFYFPAYE
:||||: :: :: ||| :  | ::|  ||| |:: |:::  :| |: ||:|:| |: |  ||:  |:|    :|   |
ITGILLALSAVATFVAGILAGFFADRNGRKPVMVFGTIIQLLGAALAIASNLPGHVNPWSTFIAFLLISFGYNFVITAGN
               60        70        80        90       100       110       120

723       753       783       813       843       873       900       930
AMTIDLTNEQNRRFVYTIGYWLVNIAVMLGSGIAGIFYDHHFFELLIVLLIISAICCFVVYFKFDET-KPQEGTFKHDKG
|| || :| :||: |: : ||  |::|:||: :   ::   |  ||::||:   :  |:  |   || ||   | | |:
AMIIDASNAENRKVVFMLDYWAQNLSVILGAALGAWLFRPAFEALLVILLLTVLVSFFLTTFVMTETFKP---TVKVDEK
              140       150       160       170       180       190          200

960       990       1020      1050      1080      1110      1140      1170
VLGTFKNYSQVLVDKAFVVYTLGAIGSSVVWLQVDNYFSVNLKQNFEVVSILGHTITGAKMLSLAVFTNTLLIVLLMTTI
    |: |  || || ::::    | :: : :| ||:: |:|  :|: ::  |  | | :||:: :    :|:||||||:
AENIFQAYKTVLQDKTYMIFMGANIATTFIIMQFDNFLPVHLSNSFKTITFWGFEIYGQRMLTIYLILACVLVVLLMTTL
       210       220       230       240       250       260       270       280

1200      1230      1260      1290      1320      1350      1380      1410
NKFIENWPLKRQLILGSLICGFGMLFNISLNTFGAILIAMTFFTFGEMIYVPASQVLRAEMMVEGKIGSYSGFLAIAQPV
|:: ::|  :: :| |||    ||:|:    ||  |:||   :|:||::| |: | | |::|   |||||:|  ||  |:
NRLTKDWSHQKGFIWGSLFMAIGMIFSFLTTTFTPIFIAGIVYTLGEIVYTPSVQTLGADLMNPEKIGSYNGVAAIKMPI
       290       300       310       320       330       340       350       360

1440      1470      1500      1530      1560      1590      1620      1650
ASVLAGAMVSLSYFTGKIGVQITLTIFMLAGLVLILYATKMKNIEIGK*NVRLY*RKIE*NNG*IYCCGNSWIGIHDICG
||:||| :||:|     ||| : | :  :  ::|:| |
ASILAGLLVSISPMIKAIGVSLVLALTEVLAIILVLVAVNRHQKTKLN
       370       380       390       400
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1083

A DNA sequence (GBSx1157) was identified in *S.agalactiae* <SEQ ID 3339> which encodes the amino acid sequence <SEQ ID 3340>. This protein is predicted to be holliday junction DNA helicase (ruvA). Analysis of this protein sequence reveals the following:

```
Possible site: 37
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.75    Transmembrane    75 -   91 (   74 -   91)

----- Final Results -----
               bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04943 GB:AP001511 holliday junction DNA helicase [Bacillus halodurans]
 Identities = 86/201 (42%), Positives = 122/201 (59%), Gaps = 6/201 (2%)

Query: 1   MYDYIKGKLSKITAKFIVVETAGLGYMIYVANPYSFSGYVNQEVTIYLHQVIRDDAHLLF 60
           M DY++G L+ I  ++ VVE  G+GY +Y  NPY F    +  +TIY  Q +R+D   L+
Sbjct: 1   MIDYLRGTLTDIDHQYAVVEVHGVGYQVYCPNPYEFEKERDSVITIYTFQYVREDVIRLY 60

Query: 61  GFHTENEKEIFLNLISVSGIGPTTALAIIAVDDNEGLVSAIDNSDIKYLTKFPKIGKKTA 120
           GF T+ ++ +F  L++VSGIGP  ALAI+A    E ++ AI+  D  +L KFP +GKKTA
Sbjct: 61  GFRTKEKRSLFEKLLNVSGIGPKGALAILATGQPEHVIQAIEEEDEAFLVKFPGVGKKTA 120

Query: 121 QQMILDLSGKFVE------ASGESATSRKVSSEQNSNLEEAMEALLALGYKATELKKVKA 174
           +Q+ILDL GK  E      +  E        ++ N  L+EAMEAL ALGY   ELKKVK
Sbjct: 121 RQIILDLKGKVDELHPGLFSQKEEQPKPHEKNDGNQALDEAMEALKALGYVEKELKKVKP 180

Query: 175 FFEGTNETVEQYIKSSLKMLM 195
             E    T + YIK +L++++
Sbjct: 181 KLEQETLTTDAYIKKALQLML 201
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3341> which encodes the amino acid sequence <SEQ ID 3342>. Analysis of this protein sequence reveals the following:

```
      Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.59    Transmembrane    75 -   91 (   74 -   91)

----- Final Results -----
               bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04943 GB:AP001511 holliday junction DNA helicase [Bacillus halodurans]
 Identities = 91/201 (45%), Positives = 128/201 (63%), Gaps = 5/201 (2%)

Query: 1   MYDYIKGQLTKITAKYIVVEANGLGYMINVANPYSFTDSVNQLVTIYLHQVIREDAHLLF 60
           M DY++G LT I  +Y VVE +G+GY +   NPY F    + ++TIY  Q +RED   L+
Sbjct: 1   MIDYLRGTLTDIDHQYAVVEVHGVGYQVYCPNPYEFEKERDSVITIYTFQYVREDVIRLY 60

Query: 61  GFHTEDEKDVFLKLISVSGIGPTTALAIVAVDDNEGLVNAIDNSDIKYLMKFPKIGKKTA 120
           GF T++++ +F KL++VSGIGP  ALAI+A    E ++ AI+  D  +L+KFP +GKKTA
Sbjct: 61  GFRTKEKRSLFEKLLNVSGIGPKGALAILATGQPEHVIQAIEEEDEAFLVKFPGVGKKTA 120

Query: 121 QQMVLDLAGKFVEA-----PQETGHTKARSNKAGNTQLDEAIEALLALGYKAKELKKIRA 175
           +Q++LDL GK  E       Q+    K      GN  LDEA+EAL ALGY  KELKK++
Sbjct: 121 RQIILDLKGKVDELHPGLFSQKEEQPKPHEKNDGNQALDEAMEALKALGYVEKELKKVKP 180

Query: 176 FFEGTSETAEQYIKSALKLLM 196
             E  + T + YIK AL+L++
Sbjct: 181 KLEQETLTTDAYIKKALQLML 201
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/197 (77%), Positives = 176/197 (88%), Gaps = 1/197 (0%)

Query: 1   MYDYIKGKLSKITAKFIVVETAGLGYMIYVANPYSFSGYVNQEVTIYLHQVIRDDAHLLF 60
```

```
           MYDYIKG+L+KITAK+IVVE  GLGYMI VANPYSF+  VNQ VTIYLHQVIR+DAHLLF
Sbjct: 1   MYDYIKGQLTKITAKYIVVEANGLGYMINVANPYSFTDSVNQLVTIYLHQVIREDAHLLF 60

Query: 61  GFHTENEKEIFLNLISVSGIGPTTALAIIAVDDNEGLVSAIDNSDIKYLTKFPKIGKKTA 120
           GFHTE+EK++FL LISVSGIGPTTALAI+AVDDNEGLV+AIDNSDIKYL KFPKIGKKTA
Sbjct: 61  GFHTEDEKDVFLKLISVSGIGPTTALAIVAVDDNEGLVNAIDNSDIKYLMKFPKIGKKTA 120

Query: 121 QQMILDLSGKFVEASGESA-TSRKVSSEQNSNLEEAMEALLALGYKATELKKVKAFFEGT 179
           QQM+LDL+GKFVEA  E+  T  + +   N+ L+EA+EALLALGYKA ELKK++AFFEGT
Sbjct: 121 QQMVLDLAGKFVEAPQETGHTKARSNKAGNTQLDEAIEALLALGYKAKELKKIRAFFEGT 180

Query: 180 NETVEQYIKSSLKMLMK 196
           +ET EQYIKS+LK+LMK
Sbjct: 181 SETAEQYIKSALKLLMK 197
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1084

A DNA sequence (GBSx1159) was identified in *S.agalactiae* <SEQ ID 3343> which encodes the amino acid sequence <SEQ ID 3344>. This protein is predicted to be DNA-3-methyladenine glycosidase I (tag). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2812(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10263> which encodes amino acid sequence <SEQ ID 10264> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76573 GB:AE000432 3-methyl-adenine DNA glycosylase I,
           constitutive [Escherichia coli K12]
 Identities = 87/176 (49%), Positives = 122/176 (68%), Gaps = 1/176 (0%)

Query: 5   MKRCSWVNLDNPLYVAYHDKEWGRAVHDDHVLFELLCLETYQSGLSWETVLNKRQEFRQV 64
           M+RC WV+ D PLY+AYHD EWG    D   LFE++CLE  Q+GLSW TVL KR+ +R
Sbjct: 1   MERCGWVSQD-PLYIAYHDNEWGVPETDSKKLFEMICLEGQQAGLSWITVLKKRENYRAC 59

Query: 65  FHHYNIEKVAAMSDADLEIILQNPRVIRHRLKLFSTRQNARSIILIQKEFGSFDRYIWSF 124
           FH ++  KVAAM + D+E ++Q+  +IRHR K+ +   NAR+ + +++    F  ++WSF
Sbjct: 60  FHQFDPVKVAAMQEEDVERLVQDAGIIRHRGKIQAIIGNARAYLQMEQNGEPFVDFVWSF 119

Query: 125 VDNKVQVNSVNNYNDVPASTTLSERLSKDLKKRGFKFVGPTCLYSFIQAAGMVNDH 180
           V+++ QV      +++P ST+ S+ LSK LKKRGFKFVG T  YSF+QA G+VNDH
Sbjct: 120 VNHQPQVTQATTLSEIPTSTSASDALSKALKKRGFKFVGTTICYSFMQACGLVNDH 175
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3345> which encodes the amino acid sequence <SEQ ID 3346>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4149(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/184 (61%), Positives = 135/184 (72%)

Query: 3   FHMKRCSWVNLDNPLYVAYHDKEWGRAVHDDHVLFELLCLETYQSGLSWETVLNKRQEFR 62
           FHMKRCSWV  DN LY  YHD EWG+ + DD   FELLCLE+YQSGLSW TVL KRQ FR
Sbjct: 2   FHMKRCSWVPKDNQLYCDYHDLEWGQPLDDDRDFFELLCLESYQSGLSWLTVLKKRQAFR 61

Query: 63  QVFHHYNIEKVAAMSDADLEIILQNPRVIRHRLKLFSTRQNARSIILIQKEFGSFDRYIW 122
            VFHHY+I  VA  +  ++   L+NP +IRH+LKL +T  NA ++  IQKEFGSF  Y+W
Sbjct: 62  TVFHHYDIASVATFTSEEMADALENPSIIRHKLKLAATVNNAIAVQKIQKEFGSFSTYLW 121

Query: 123 SFVDNKVQVNSVNNYNDVPASTTLSERLSKDLKKRGFKFVGPTCLYSFIQAAGMVNDHEN 182
           +FV  K   N VN  N VPA T LS RL+KDLKKRGFKF+GPT +YSF+QA+G+VNDHE
Sbjct: 122 NFVGGKPINNLVNQENLVPAQTELSIRLAKDLKKRGFKFLGPTTVYSFMQASGLVNDHEE 181

Query: 183 ICDF 186
            C F
Sbjct: 182 ACVF 185
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1085

A DNA sequence (GBSx1160) was identified in *S.agalactiae* <SEQ ID 3347> which encodes the amino acid sequence <SEQ ID 3348>. This protein is predicted to be competence-damage inducible protein (cinA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10261> which encodes amino acid sequence <SEQ ID 10262> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA84071 GB:Z34303 CinA protein [Streptococcus pneumoniae]
 Identities = 194/297 (65%), Positives = 236/297 (79%), Gaps = 1/297 (0%)

Query: 1   MVEGSIPLQNLTGLAVGGIVTSKGVQYMVLPGPPSELKPMVMEQVVPILSNNGTKLYSRV 60
           +VEG+IPL N TGLAVGG +   GV Y+VLPGPPSELKPMV+ Q++P L   G+KLYSRV
Sbjct: 121 IVEGAIPLPNETGLAVGGKLEVDGVTYVVLPGPPSELKPMVLNQLLPKLMT-GSKLYSRV 179

Query: 61  LRFFGIGESQLVTILEDIIKNQTDPTIAPYAKVGEVTLRLSTKAENQDEADFKLDSLEKE 120
           LRFFGIGESQLVTIL D+I NQ DPT+APYAK GEVTLRLSTKA +Q+EA+  LD LE +
Sbjct: 180 LRFFGIGESQLVTILADLIDNQIDPTLAPYAKTGEVTLRLSTKASSQEEANQALDILENQ 239

Query: 121 ILALKTLDNRKLKDLLYGYGDNNSMARTVLELLKVQNKTITAAESLTAGLFQSQLAEFSG 180
           IL  +T +   L+D  YGYG+  S+A  V+E LK Q KTI AAESLTAGLFQ+ +A FSG
Sbjct: 240 ILDCQTFEGISLRDFCYGYGEETSLASIVVEELKRQGKTIAAAESLTAGLFQATVANFSG 299

Query: 181 ASQVFNGGFTTYSMEAKSQLLGIPKKKLQEYGVVSHFTAEAMAQQARQLLKADFGIGLTG 240
            S +F GGF TYS+E KS++L IP K L+E+GVVS FTA+ MA+QAR   ++DFGI LTG
Sbjct: 300 VSSIFEGGFVTYSLEEKSRMLDIPAKNLEEHGVVSEFTAQKMAEQARSKTQSDFGISLTG 359

Query: 241 VAGPDELEGYPAGTVFIGIATPEGVSSIKVSIGGKSRSDVRHISTLHAFDLVRRALL 297
           VAGPD LEG+P GTVFIG+A  +G   IKV+IGG+SR+DVRHI+ +HAF+LVR+ALL
Sbjct: 360 VAGPDSLEGHPVGTVFIGLAQDQGTEVIKVNIGGRSRADVRHIAVMHAFNLVRKALL 416
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3349> which encodes the amino acid sequence <SEQ ID 3350>. Analysis of this protein sequence reveals the following:

```
        Possible site: 22

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.91   Transmembrane  134 - 150 ( 134 - 150)

----- Final Results -----
               bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA84071 GB:Z34303 CinA protein [Streptococcus pneumoniae]
 Identities = 286/417 (68%), Positives = 336/417 (79%), Gaps = 1/417 (0%)

Query: 1   MKAELIAVGTEILTGQIVNTNAQFLSEKMAELGIDVYFQTAVGDNEERLLSVITTASQRS 60
           MKAE+IAVGTEILTGQIVNTNAQFLSEK+AE+G+DVYFQTAVGDNE RLLS++  ASQRS
Sbjct: 1   MKAEIIAVGTEILTGQIVNTNAQFLSEKLAEIGVDVYFQTAVGDNEVRLLSLLEIASQRS 60

Query: 61  NLVILCGGLGPTKDDLTKQTLAKYLRKDLVYDEQACQKLDDFFAKRKPSSRTPNNERQAQ 120
           +LVIL GGLG T+DDLTKQTLAK+L K LV+D QA +KLD FFA R   +RTPNNERQAQ
Sbjct: 61  SLVILTGGLGATEDDLTKQTLAKFLGKALVFDPQAQEKLDIFFALRPDYARTPNNERQAQ 120

Query: 121 VIEGSIPLPNKTGLAVGGFITVDGISYVVLPGPPSELKPMVNEELVPLLSKQYSTLYSKV 180
           ++EG+IPLPN+TGLAVGG + VDG++YVVLPGPPSELKPMV  +L+P L    S LYS+V
Sbjct: 121 IVEGAIPLPNETGLAVGGKLEVDGVTYVVLPGPPSELKPMVLNQLLPKLMTG-SKLYSRV 179

Query: 181 LRFFGIGESQLVTVLSDFIENQTDPTIAPYAKTGEVTLRLSTKTENQALADKKLGQLEAQ 240
           LRFFGIGESQLVT+L+D I+NQ DPT+APYAKTGEVTLRLSTK  +Q  A++ L  LE Q
Sbjct: 180 LRFFGIGESQLVTILADLIDNQIDPTLAPYAKTGEVTLRLSTKASSQEEANQALDILENQ 239

Query: 241 LLSRKTLEGQPLADVFYGYGEDNSLARETFELLVKYDKTITAAESLTAGLFQSTLASFPG 300
           +L  +T EG  L D  YGYGE+ SLA    E L +  KTI AAESLTAGLFQ+T+A+F G
Sbjct: 240 ILDCQTFEGISLRDFCYGYGEETSLASIVVEELKRQGKTIAAAESLTAGLFQATVANFSG 299

Query: 301 ASQVFNGGFVTYSMEEKAKMLGLPLEELKSHGVVSAYTAEGMAEQARLLTGADIGVSLTG 360
            S +F GGFVTYS+EEK++ML +P + L+ HGVVS +TA+ MAEQAR  T +D G+SLTG
Sbjct: 300 VSSIFEGGFVTYSLEEKSRMLDIPAKNLEEHGVVSEFTAQKMAEQARSKTQSDFGISLTG 359

Query: 361 VAGPDMLEEQPAGTVFIGLATQNKVESIKVLISGRSRLDVRYIATLHAFNMVRKTLL 417
           VAGPD LE  P GTVFIGLA     E IKV I GRSR DVR+IA +HAFN+VRK LL
Sbjct: 360 VAGPDSLEGHPVGTVFIGLAQDQGTEVIKVNIGGRSRADVRHIAVMHAFNLVRKALL 416
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/299 (67%), Positives = 242/299 (80%)

Query: 1   MVEGSIPLQNLTGLAVGGIVTSKGVQYMVLPGPPSELKPMVMEQVVPILSNNGTKLYSRV 60
           ++EGSIPL N TGLAVGG +T  G+ Y+VLPGPPSELKPMV E++VP+LS    + LYS+V
Sbjct: 121 VIEGSIPLPNKTGLAVGGFITVDGISYVVLPGPPSELKPMVNEELVPLLSKQYSTLYSKV 180

Query: 61  LRFFGIGESQLVTILEDIIKNQTDPTIAPYAKVGEVTLRLSTKAENQDEADFKLDSLEKE 120
           LRFFGIGESQLVT+L D I+NQTDPTIAPYAK GEVTLRLSTK ENQ  AD KL  LE +
Sbjct: 181 LRFFGIGESQLVTVLSDFIENQTDPTIAPYAKTGEVTLRLSTKTENQALADKKLGQLEAQ 240

Query: 121 ILALKTLDNRKLKDLLYGYGDNNSMARTVLELLKVQNKTITAAESLTAGLFQSQLAEFSG 180
           +L+ KTL+  + L D+ YGYG++NS+AR   ELL   +KTITAAESLTAGLFQS LA F G
Sbjct: 241 LLSRKTLEGQPLADVFYGYGEDNSLARETFELLVKYDKTITAAESLTAGLFQSTLASFPG 300

Query: 181 ASQVFNGGFTTYSMEAKSQLLGIPKKKLQEYGVVSHFTAEAMAQQARQLLKADFGIGLTG 240
           ASQVFNGGF TYSME K+++LG+P ++L+ +GVVS +TAE MA+QAR L  AD G+ LTG
Sbjct: 301 ASQVFNGGFVTYSMEEKAKMLGLPLEELKSHGVVSAYTAEGMAEQARLLTGADIGVSLTG 360
```

```
Query: 241 VAGPDELEGYPAGTVFIGIATPEGVSSIKVSIGGKSRSDVRHISTLHAFDLVRRALLKI 299
           VAGPD LE  PAGTVFIG+AT   V SIKV I G+SR DVR+I+TLHAF++VR+ LLK+
Sbjct: 361 VAGPDMLEEQPAGTVFIGLATQNKVESIKVLISGRSRLDVRYIATLHAFNMVRKTLLKL 419
```

SEQ ID 3348 (GBS646) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 2-4; MW 61.6kDa), in Figure 134 (lane 3; MW 57.5kDa + lanes 2 & 4; MW 27kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 5-7; MW 36.6kDa) and in Figure 178 (lane 5; MW 37kDa).

GBS646-His was purified as shown in Figure 229, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1086

A DNA sequence (GBSx1161) was identified in *S.agalactiae* <SEQ ID 3351> which encodes the amino acid sequence <SEQ ID 3352>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  148 - 164 ( 148 - 164)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3353> which encodes the amino acid sequence <SEQ ID 3354>. Analysis of this protein sequence reveals the following:

```
     Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  148 - 164 ( 148 - 164)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD04860 GB:AF069745 RecA protein [Streptococcus parasanguinis]
 Identities = 333/381 (87%), Positives = 356/381 (93%), Gaps = 3/381 (0%)

Query: 1   LAKKLKKNEEITKKFGDERRKALDDALKNIEKDFGKGAVMRLGERAEQKVQVMSSGSLAL 60
            +AKK KK ++ITKKFGDER KAL+DALK IEKDFGKG++MRLGERAEQKVQVMSSGSLAL
Sbjct: 1   MAKKQKKLDDITKKFGDEREKALNDALKLIEKDFGKGSIMRLGERAEQKVQVMSSGSLAL 60

Query: 61  DIALGAGGYPKGRIIEIYGPESSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL 120
            DIALGAGGYPKGRIIEIYGPESSGKTTVALHAVAQAQKEGGIAAFIDAEHALDP+YAAAL
Sbjct: 61  DIALGAGGYPKGRIIEIYGPESSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPSYAAAL 120

Query: 121 GVNIDELLLSQPDSGEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGDIGDSHVGLQ 180
            GVNIDELLLSQPDSGEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGDIGDSHVGLQ
Sbjct: 121 GVNIDELLLSQPDSGEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGDIGDSHVGLQ 180

Query: 181 ARMMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240
            ARMMSQAMRKL ASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG
Sbjct: 181 ARMMSQAMRKLGASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240

Query: 241 TTQIKGTGDQKDSSIGKETKIKVVKNKVAPPFKVAEVEIMYGEGISRTGELVKIASDLDI 300
             TQIKGTGDQKD+++GKETKIKVVKNKVAPPFK A VEIMYGEGISRTGELVKIA+DLDI
```

```
Sbjct: 241 NTQIKGTGDQKDTNVGKETKIKVVKNKVAPPFKEAMVEIMYGEGISRTGELVKIATDLDI 300

Query: 301 IQKAGAWFSYNGEKIGQGSENAKRYLADHPELFDEIDLKVRVKFGLLEESEEESAMAVAS 360
           IQKAGAW+SYNGEKIGQGSENAK++LADHPE+FDEID KVRV FGL+E+ E   ++
Sbjct: 301 IQKAGAWYSYNGEKIGQGSENAKKFLADHPEIFDEIDHKVRVHFGLIEKDEAVKSLDKTE 360

Query: 361 EE---TDDLALDLDNGIEIED 378
           E     +++ LDLD+ IEIED
Sbjct: 361 EAAPVVEEVTLDLDDAIEIED 381
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 339/379 (89%), Positives = 356/379 (93%), Gaps = 1/379 (0%)

Query: 1   MAKKTKKAEEITKKFGDERRKALDDALKNIEKDFGKGAVMRLGERAEQKVQVMSSGSLAL 60
           +AKK KK EEITKKFGDERRKALDDALKNIEKDFGKGAVMRLGERAEQKVQVMSSGSLAL
Sbjct: 1   LAKKLKKNEEITKKFGDERRKALDDALKNIEKDFGKGAVMRLGERAEQKVQVMSSGSLAL 60

Query: 61  DIALGAGGYPKGRIVEIYGPESSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL 120
           DIALGAGGYPKGRI+EIYGPESSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL
Sbjct: 61  DIALGAGGYPKGRIIEIYGPESSGKTTVALHAVAQAQKEGGIAAFIDAEHALDPAYAAAL 120

Query: 121 GVNIDELLLSQPDSGEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGDIGDSHVGLQ 180
           GVNIDELLLSQPDSGEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGDIGDSHVGLQ
Sbjct: 121 GVNIDELLLSQPDSGEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGDIGDSHVGLQ 180

Query: 181 ARMMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYSSVRLDVRG 240
           ARMMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFY+SVRLDVRG
Sbjct: 181 ARMMSQAMRKLSASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240

Query: 241 NTQIKGTGEHKDHNVGKETKIKVVKNKVAPPFREAFVEIMYGEGISRTGELIKIASDLDI 300
            TQIKGTG+ KD ++GKETKIKVVKNKVAPPF+ A VEIMYGEGISRTGEL+KIASDLDI
Sbjct: 241 TTQIKGTGDQKDSSIGKETKIKVVKNKVAPPFKVAEVEIMYGEGISRTGELVKIASDLDI 300

Query: 301 IQKAGAWYSYNGEKIGQGSENAKKYLADNPAIFDEIDHKVRVHFGMTEDDSPVQSELVEE 360
           IQKAGAW+SYNGEKIGQGSENAK+YLAD+P +FDEID KVRV FG+ E +S  +S +
Sbjct: 301 IQKAGAWFSYNGEKIGQGSENAKRYLADHPELFDEIDLKVRVKFGLLE-ESEEESAMAVA 359

Query: 361 KNEADDLVLDLDNAIEIEE 379
             E DDL LDLDN IEIE+
Sbjct: 360 SEETDDLALDLDNGIEIED 378
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1087

A DNA sequence (GBSx1162) was identified in *S.agalactiae* <SEQ ID 3355> which encodes the amino acid sequence <SEQ ID 3356>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2344(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10259> which encodes amino acid sequence <SEQ ID 10260> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG37358 GB:AF028804 NrpR [Lactococcus lactis subsp. cremoris]
 Identities = 69/132 (52%), Positives = 102/132 (77%)
```

```
Query: 5   MIKIYTISSCTSCKKAKTWLNAHQLPYKEQNLGKESLTRDEILEILTKTESGIESIVSSK 64
           MI IYT  SCTSCKKAKTWL+ H +P+ E+NL  + L+  EI +IL K + G+E ++SS+
Sbjct: 1   MITIYTAPSCTSCKKAKTWLSYHHIPFNERNLIADPLSTTEISQILQKCDDGVEGLISSR 60

Query: 65  NRYAKALNCNIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLPRSIRN 124
           NR+ K L  + E++S+++ I +I ENP+I++ PI++D+KRL VGY E++IRAFLPR++R
Sbjct: 61  NRFVKTLGVDFEDISLSQAIKIISENPQIMRRPIIMDEKRLHVGYNEEEIRAFLPRTVRV 120

Query: 125 VENAEARLRAAL 136
           +EN  ARLR+A+
Sbjct: 121 LENGGARLRSAI 132
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3357> which encodes the amino acid sequence <SEQ ID 3358>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2569(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 117/132 (88%), Positives = 128/132 (96%)

Query: 5   MIKIYTISSCTSCKKAKTWLNAHQLPYKEQNLGKESLTRDEILEILTKTESGIESIVSSK 64
           MIKIYTISSCTSCKKAKTWLNAH+L YKEQNLGKE LT++EIL IL+KTE+G+ESIVSSK
Sbjct: 1   MIKIYTISSCTSCKKAKTWLNAHKLAYKEQNLGKEPLTKEEILAILSKTENGVESIVSSK 60

Query: 65  NRYAKALNCNIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLPRSIRN 124
           NRYAKAL+C+IEELSV+EVIDLIQ+NPRILKSPILIDDKRLQVGYKEDDIRAFLPRSIRN
Sbjct: 61  NRYAKALDCDIEELSVSEVIDLIQDNPRILKSPILIDDKRLQVGYKEDDIRAFLPRSIRN 120

Query: 125 VENAEARLRAAL 136
           +EN EARLRAAL
Sbjct: 121 IENTEARLRAAL 132
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1088

A DNA sequence (GBSx1163) was identified in *S.agalactiae* <SEQ ID 3359> which encodes the amino acid sequence <SEQ ID 3360>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04987 GB:AP001511 unknown [Bacillus halodurans]
 Identities = 49/82 (59%), Positives = 64/82 (77%), Gaps = 1/82 (1%)

Query: 1   MGFTDETVRFRLDDSN-KVEISETLTAVYRSLEEKGYNPINQIVGYVLSGDPAYVPRYND 59
           M   D T++F +++    V++ E L +VY +LEEKGYNPINQIVGY+LSGDPAY+PR+ D
Sbjct: 1   MSSMDNTMKFNVNEEPVSVDVQEVLMSVYEALEEKGYNPINQIVGYLLSGDPAYIPRHKD 60
```

```
Query: 60 ARNQIRKYERDEIVEELVRYYL 81
          AR  IRK ERDE++EELV+ YL
Sbjct: 61 ARTLIRKLERDELIEELVKSYL 82
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3361> which encodes the amino acid sequence <SEQ ID 3362>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 80/88 (90%), Positives = 85/88 (95%)

Query: 1  MGFTDETVRFRLDDSNKVEISETLTAVYRSLEEKGYNPINQIVGYVLSGDPAYVPRYNDA 60
          MGFTDETVRF+LDD +K +ISETLTAVY SL+EKGYNPINQIVGYVLSGDPAYVPRYNDA
Sbjct: 1  MGFTDETVRFKLDDGDKRQISETLTAVYHSLDEKGYNPINQIVGYVLSGDPAYVPRYNDA 60

Query: 61 RNQIRKYERDEIVEELVRYYLQGNGIDL 88
          RNQIRKYERDEIVEELVRYYLQGNGID+
Sbjct: 61 RNQIRKYERDEIVEELVRYYLQGNGIDV 88
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1089

A DNA sequence (GBSx1164) was identified in *S.agalactiae* <SEQ ID 3363> which encodes the amino acid sequence <SEQ ID 3364>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1575(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10257> which encodes amino acid sequence <SEQ ID 10258> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14698 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 82/138 (59%), Positives = 109/138 (78%), Gaps = 1/138 (0%)

Query: 1   MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEESGNFGFDRLAELVKEYKVDKFVVG 60
           MRI+GLD+G+KT+GVA+SD +G+TAQG+E IKI+E  G++G  RL+EL+K+Y +DK V+G
Sbjct: 1   MRILGLDLGTKTLGVALSDEMGWTAQGIETIKINEAEGDYGLSRLSELIKDYTIDKIVLG 60

Query: 61  LPKNMNNTSGPRVEASQAYGDKITELFNLPVEYQDERLTTVQAERMLVEQADISRGKRKK 120
            PKNMN T GPR EASQ +   +   +N+PV   DERLTT+ AE+ML+  AD+SR KRKK
Sbjct: 61  FPKNMNGTVGPRGEASQTFAKVLETTYNVPVVLWDERLTTMAAEKMLI-AADVSRQKRKK 119

Query: 121 VIDKLAAQLILQNYLDRM 138
           VIDK+AA +ILQ YLD +
Sbjct: 120 VIDKMAAVMILQGYLDSL 137
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3365> which encodes the amino acid sequence <SEQ ID 3366>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1575(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/139 (82%), Positives = 126/139 (90%)

Query: 1   MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEESGNFGFDRLAELVKEYKVDKFVVG 60
           MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEE   FGF RL ELVK+Y+V++FV+G
Sbjct: 1   MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEEKAEFGFTRLEELVKQYQVEQFVIG 60

Query: 61  LPKNMNNTSGPRVEASQAYGDKITELFNLPVEYQDERLTTVQAERMLVEQADISRGKRKK 120
           LPKNMNNT+GPRV+AS  YG+ I  LF LPV YQDERLTTV+A+RML+EQADISRGKRKK
Sbjct: 61  LPKNMNNTNGPRVDASITYGNHIEHLFGLPVHYQDERLTTVEAKRMLIEQADISRGKRKK 120

Query: 121 VIDKLAAQLILQNYLDRMF 139
           VIDKLAAQLILQNYL+R F
Sbjct: 121 VIDKLAAQLILQNYLNRNF 139
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1090

A DNA sequence (GBSx1165) was identified in *S.agalactiae* <SEQ ID 3367> which encodes the amino acid sequence <SEQ ID 3368>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2631(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14697 GB:Z99118 yrzB [Bacillus subtilis]
 Identities = 50/94 (53%), Positives = 65/94 (68%), Gaps = 5/94 (5%)

Query: 12  EHQHEVITLVDENGNETLFEILLTIDGREEFGKNYVLLVPAGAEEDEQGEIEIQAYSFTE 71
           EH   + IT+VD+ GNE L E+L T +  EEFGK+YVL  P  +++DE  E+EI A SFT
Sbjct: 2   EHGEKNITIVDDQGNEQLCEVLFTFEN-EEFGKSYVLYYPIESKDDE--EVEILASSFTP 58

Query: 72  NADGTEGDLQPIPEDSDAEWDMIEEVFNSFLDEE 105
           N DG  G+L PI  ++D EWDMIEE  N+FL +E
Sbjct: 59  NEDGENGELFPI--ETDEEWDMIEETLNTFLADE 90
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3369> which encodes the amino acid sequence <SEQ ID 3370>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3170(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 90/98 (91%), Positives = 94/98 (95%)

Query: 7    HDHNHEHQHEVITLVDENGNETLFEILLTIDGREEFGKNYVLLVPAGAEEDEQGEIEIQA 66
             H+H ++HQHEVITLVDE GNETLFEILLTIDGREEFGKNYVLLVPAG+EEDE GEIEIQA
Sbjct: 3    HNHENDHQHEVITLVDEQGNETLFEILLTIDGREEFGKNYVLLVPAGSEEDESGEIEIQA 62

Query: 67   YSFTENADGTEGDLQPIPEDSDAEWDMIEEVFNSFLDE 104
             YSFTEN DGTEGDLQPIPEDSDAEWDMIEEVFNSFLDE
Sbjct: 63   YSFTENEDGTEGDLQPIPEDSDAEWDMIEEVFNSFLDE 100
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1091

A DNA sequence (GBSx1166) was identified in *S.agalactiae* <SEQ ID 3371> which encodes the amino acid sequence <SEQ ID 3372>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1092

A DNA sequence (GBSx1167) was identified in *S.agalactiae* <SEQ ID 3373> which encodes the amino acid sequence <SEQ ID 3374>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.18   Transmembrane  314 - 330 ( 308 - 334)
   INTEGRAL    Likelihood = -6.21   Transmembrane  279 - 295 ( 274 - 300)
   INTEGRAL    Likelihood = -6.10   Transmembrane  136 - 152 ( 135 - 157)
   INTEGRAL    Likelihood = -5.31   Transmembrane  232 - 248 ( 226 - 253)
   INTEGRAL    Likelihood = -4.73   Transmembrane  163 - 179 ( 162 - 180)
   INTEGRAL    Likelihood = -3.13   Transmembrane   95 - 111 (  94 - 119)
   INTEGRAL    Likelihood = -3.03   Transmembrane  386 - 402 ( 386 - 405)
   INTEGRAL    Likelihood = -2.18   Transmembrane  204 - 220 ( 204 - 221)
   INTEGRAL    Likelihood = -2.13   Transmembrane   40 -  56 (  40 -  57)
   INTEGRAL    Likelihood = -1.70   Transmembrane  186 - 202 ( 182 - 202)

----- Final Results -----
               bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10255> which encodes amino acid sequence <SEQ ID 10256> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3375> which encodes the amino acid sequence <SEQ ID 3376>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.38   Transmembrane  315 - 331 ( 311 - 333)
   INTEGRAL    Likelihood = -6.48   Transmembrane   40 -  56 (  37 -  61)
   INTEGRAL    Likelihood = -6.10   Transmembrane  278 - 294 ( 274 - 298)
   INTEGRAL    Likelihood = -5.57   Transmembrane  392 - 408 ( 387 - 410)
   INTEGRAL    Likelihood = -3.98   Transmembrane  186 - 202 ( 184 - 208)
   INTEGRAL    Likelihood = -3.93   Transmembrane  339 - 355 ( 338 - 356)
   INTEGRAL    Likelihood = -2.97   Transmembrane  235 - 251 ( 228 - 253)
   INTEGRAL    Likelihood = -2.44   Transmembrane  166 - 182 ( 166 - 182)
   INTEGRAL    Likelihood = -2.23   Transmembrane  106 - 122 ( 106 - 125)
   INTEGRAL    Likelihood = -1.81   Transmembrane   83 -  99 (  83 - 101)

----- Final Results -----
               bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9179> which encodes the amino acid sequence <SEQ ID 9180>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 13
>>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood = -7.38   Transmembrane  243 - 259 ( 239 - 261)
     INTEGRAL    Likelihood = -6.10   Transmembrane  206 - 222 ( 202 - 226)
     INTEGRAL    Likelihood = -5.57   Transmembrane  320 - 336 ( 315 - 338)
     INTEGRAL    Likelihood = -3.98   Transmembrane  114 - 130 ( 112 - 136)
     INTEGRAL    Likelihood = -3.93   Transmembrane  267 - 283 ( 266 - 284)
     INTEGRAL    Likelihood = -2.97   Transmembrane  163 - 179 ( 156 - 181)
     INTEGRAL    Likelihood = -2.44   Transmembrane   94 - 110 (  94 - 110)
     INTEGRAL    Likelihood = -2.23   Transmembrane   34 -  50 (  34 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.395(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 200/480 (41%), Positives = 310/480 (63%), Gaps = 1/480 (0%)

Query: 40  ILLYSVLSTLLAIANPLLTYFANGLQTQNLYTGLMMTKGQIPYSDVFATGGFLYYVTIAL 99
           +L +S++ + L IA P LT  ANGLQ+QNLY G+M+TKGQ+PYS  F TGG  Y+V IAL
Sbjct: 40  LLFFSIIISSLTIAVPFLTDAANGLQSQNLYIGMMLTKGQLPYSAAFTTGGLFYFVIIAL 99

Query: 100 SYLLGSSIWLLIVQFIAYYVSGIYFYKLVYYVAQSEIVSIGMTLIFYIMNIVLGFGGMYP 159
           SY LGS++WL+ VQ   +Y+SG+Y YKL+ Y+   + V++  ++ +Y++++ LGFGG+YP
Sbjct: 100 SYYLGSTLWLVFVQVFCFYLSGLYLYKLINYMTGFQKVALTFSISYYLLSVSLGFGGLYP 159

Query: 160 IQWALPFMLISLWFLIKFCVDNIVDEAFIFYGILAAFSLFIDPQTLIFWLCSFVLLTATN 219
            Q A+PF+LIS WFL K+    + DEAFI +G + A ++ IDP TLIFW  + V + + N
Sbjct: 160 TQLAMPFILISAWFLTKYFACLVKDEAFILFGFVGALAMLIDPSTLIFWSFACVTVFSYN 219

Query: 220 IKQKQSLRGFYQFLCVVFGMILIAYTVGYFMFNLQIISSYIDKAIFYPFTYFARTNHSFL 279
           I QK   RGFYQ L  +FGMIL+ YT GYF+ NLQ+++ Y+ + + YPFT+F   N S L
Sbjct: 220 ISQKHLARGFYQLLASIFGMILVFYTAGYFILNLQVLNPYLSQTMIYPFTFFKSGNLSLL 279

Query: 280 LSLAIQIVVLLGSGCLFGLWDFIQNRKKASYQIGLNFIACIFIIYAIMAIFSRDFNLYHF 339
             LAIQ+   LG G L G+ + I+ K  S ++        + +   ++AIFS+D+  YH
Sbjct: 280 FGLAIQLFFALGLGLLTGMENVIRRFKNNSDRVVKWLFVMVILESILVAIFSQDYRPYHL 339
```

```
Query: 340 LPALPFGLLLTSNKITILYQKVIDRRSHRRQY-FSGKSLIVDLFVKKTYYLPLLLVSLSI 398
           LP LPFGL+LT+  +   Y   + + SHRR++  +G   ++ +++K+ +YLP+L+V   +
Sbjct: 340 LPLLPFGLILTAIPVGYQYGIGLGQSSHRRRHGKNGVGRVMMIYLKRHFYLPILIVGTIL 399

Query: 399 GLLVYNTYQNVTLSKERRDISHYLTTKIDRDGKIYVWDKVASIYSQTRLKSASQFVLPHI 458
               Y    ++ L++ER  I+ YL  K+++   IYVWD  + IY  ++ KS SQF  P I
Sbjct: 400 ICSTYCFISSIPLNQERDHIASYLEQKLNKTQSIYVWDDTSKIYLDSKAKSVSQFSSPDI 459

Query: 459 NTAQKNNEKILKDELLQHGAKYFILNKNEKLPNELKSDIKKHYQEVPLSNITHFVLYRFK 518
           NT ++++ KIL+DELL++ A Y ++N+ + LP  ++  +  +Y+         F++Y+ K
Sbjct: 460 NTQKESHRKILEDELLENKAAYIVVNRYKNLPKIIQKVLSTNYKVDKQITTKSFIVYQKK 519
```

A related GBS gene <SEQ ID 8727> and protein <SEQ ID 8728> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
SRCFLG: 0
McG: Length  of  UR:   34
     Peak Value of UR:   2.23
     Net Charge of CR: 0
McG: Discrim Score:      7.72
GvH: Signal Score (-7.5): -2.21
     Possible site: 60
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 61
ALOM program   count: 5 value:  -9.18 threshold:  0.0
   INTEGRAL    Likelihood = -9.18   Transmembrane  174 - 190 ( 168 - 194)
   INTEGRAL    Likelihood = -6.21   Transmembrane  139 - 155 ( 134 - 160)
   INTEGRAL    Likelihood = -5.31   Transmembrane   92 - 108 (  86 - 113)
   INTEGRAL    Likelihood = -3.03   Transmembrane  246 - 262 ( 246 - 265)
   INTEGRAL    Likelihood = -2.18   Transmembrane   64 -  80 (  64 -  81)
   PERIPHERAL  Likelihood =  3.29     194
 modified ALOM score:   2.34
icm1 HYPID: 7  CFP: 0.467

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02392(331 - 978 of 1764)
EGAD|43696|MJ1079(2 - 379 of 397) conserved hypothetical protein {Methanococcus jannaschii}
OMNI|MJ1079  conserved  hypothetical  protein  GP|1591727|gb|AAB99076.1||U67550  conserved
hypothetical  protein  {Methanococcus  jannaschii}  PIR|F64434|F64434  hypothetical  protein
MJ1079 - Methanococcus jannaschii
%Match = 3.1
%Identity = 25.6  %Similarity = 50.7
Matches = 57  Mismatches = 100  Conservative Sub.s = 56

174       204       234       264       294       324       354
*LLLANI*LSVHPTSFFTXXXN*LXXSSIWLLIVQFIAYYVSGIYFYKLVYYVAQSEIVSIGMTLIFYIMNIVLG-----
                                                              : |:: |: |: |
                                                             MLNLLYLILGIICGTITGL
                                                                     10

426       447       477       507       537       567       597
FGGMYPIQW-ALPFMLISLWFL---IKFCVDNIVDEAFIFYGILAAFSLFIDPQTLIFWLCSFVLLTATNIKQKQSLRGF
| |::|    || |:::  : |   | | :  ::   || : | :||    | :| :  |    |    |  :   | ||
FPGIHPNNIVALSFLILPYFGLDNYIPFLIGLVITHYFINF-IPSAFLGVPDDETAVSALPMHKLTLNGNGYEAIVLAGF
         30        40        50        60         70        80        90

627       657       687       717       747       774
```

```
YQFLCVVFGMILIAYTVGYFMFNLQIISSYIDKAIFYPFTYFARTNHSFLLSLAI-QIVVLLGSGC--------------
  :| ||| :::  : :  : |:::     |   || ||   |   :  : : :: :::|:: ||
GSYLGVVFSILISLFLMSILHFDVRAFYCSI--KIFIPFILIAFILYQIFTAKSVWEVLVIFLSGIFGIAVLYCSEAFNI
         110       120         130       140       150       160       170

798       828                                               846       876
--------LFGLWDFIQNRKKASYQ--------------------~~~~--------------------IGLNFIACIFI
        :||:  :| | |    :                                            : : |::  ||
TLTAIFTGMFGIPLLINNLKTYKIKSQMMAFPDFELKFLKSSFFA~~~~TIAIIILLNLSKYILLFIRKVNFKFLSLFFI
           190       200       210       220              320       330

906                 948       978      1008      1038      1068      1098
IYAIMAIFSRDFN---LYH---FLPALPFGLLLTSNKITILYQKVIDRRSHRRQYFSGKSLIVDLFVKKTYYLPLLLVSL
|:  : :    :|   :||   :| |:  |||   :   :
IFCSLVVIIGSYNTYLIYHIIVYLTAIYIGLLAVKSNTNLSNMMNVLIFPTILYFLRG
        350       360       370       380       390
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1093

A DNA sequence (GBSx1168) was identified in *S.agalactiae* <SEQ ID 3377> which encodes the amino acid sequence <SEQ ID 3378>. This protein is predicted to be anaerobic ribonucleotide reductase (nrdD). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3722(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10253> which encodes amino acid sequence <SEQ ID 10254> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD00215 GB:U73336 anaerobic ribonucleotide reductase
           [Lactococcus lactis subsp. cremoris]
 Identities = 539/725 (74%), Positives = 616/725 (84%), Gaps = 7/725 (0%)

Query: 10  MTESDIKVIKRDGRLVSFDKYKIYTALLKASNKVIKMSPLVEAKLEMIADHVIAEIYNRF 69
           +T  +I VIKRDGR V F+  KI+ AL KA+ KV      V   L  + D V++EI++RF
Sbjct: 10  VTLEEINVIKRDGRSVKFNSEKIFDALTKAAKKVELTDKSV---LSELTDRVVSEIFSRF 66

Query: 70  KDNIKIYEIQNIVEHKLLEANEYAIAQEYINYRTQRDFERSQATDINFSIGKLINKDQTV 129
            +N+KIYEIQ+IVE +LLE+ E A+A+EYI+YR  RD  R++ATDINF+I KLIN+DQTV
Sbjct: 67  SENVKIYEIQSIVEQELLESGETALAEEYISYRANRDLARTKATDINFTIEKLINRDQTV 126

Query: 130 VNENANKDSDVFNTQRDLTAGIVGKSIGLKMLPSHVANAHQKGDIHYHDLDYSPYTPMTN 189
           VNENANKDS+VFNTQRDLTAG V K+IGLK+LP HVANAHQKGDIHYHDLDYSP+T M N
Sbjct: 127 VNENANKDSNVFNTQRDLTAGAVSKAIGLKLLPPHVANAHQKGDIHYHDLDYSPFTTMAN 186

Query: 190 CCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTADRIDEFLAP 249
           CCLIDFK M  NGFK+GNA+V+SPKSIQTATAQ SQIIANVASSQYGGC+ DR DE LAP
Sbjct: 187 CCLIDFKNMFENGFKLGNAQVDSPKSIQTATAQASQIIANVASSQYGGCSFDRADEVLAP 246

Query: 250 YAQLNYQKHLKDAKEWVIED-KQEDYARAKTQKDIYDAMQSLEYEINTLFTSNGQTPFTS 308
           YA+LNYQKHLKDA++W+  D K+E YAR KT KDIYDAMQSLEYEINTLFTSNGQTPF +
Sbjct: 247 YAKLNYQKHLKDAQKWIDGDEKREAYAREKTAKDIYDAMQSLEYEINTLFTSNGQTPFVT 306

Query: 309 LGFGLGTNWFEREIQKAILKIRIQGLGSEHRTAIFPKLIFTLKKGLNLEEDSPNYDIKQL 368
```

```
           +GFGLG +W+ REIQKAILK+RI GLGSEHRTAIFPKLIFTLK+GLNLE  +PNYDIK+L
Sbjct: 307 VGFGLGDDWYAREIQKAILKVRIGGLGSEHRTAIFPKLIFTLKRGLNLEVGTPNYDIKEL 366

Query: 369 ALECATKRMYPDVLSYDKIIDLTGSFKAPMGCRSFLQGWRDANGQDVTSGRMNLGVVTVN 428
           ALEC+TKRMYPD+LSYDKI++LTGSFKA MGCRSFLQGW+DANG DVT+GR NLGVVTVN
Sbjct: 367 ALECSTKRMYPDILSYDKIVELTGSFKASMGCRSFLQGWKDANGNDVTAGRNNLGVVTVN 426

Query: 429 LPRVAMESNGDMDKFWEIFNERMSIARDALVYRVERVKEAIPANAPILYQYGAFGERLGK 488
           LPR+A+E+ G+ +KFWEIFNER+ IA DAL +RVER KEA P NAPIL+  GA G RL
Sbjct: 427 LPRIALEAAGNKEKFWEIFNERVEIAHDALAFRVERAKEAQPKNAPILFMNGALG-RLDS 485

Query: 489 YDNVDRLFNHRRATVSLGYIGLYEVASVFYGGDWEDNHQAKAFTVDIVRKMKQLCADWSD 548
             +VD L+N+ RATVSLGYIGLYEVA+ FYG  WE N +AKAFT++IV++M + C DWS
Sbjct: 486 EGSVDDLYNNERATVSLGYIGLYEVATTFYGPTWESNPEAKAFTIEIVKRMHEDCEDWSK 545

Query: 549 EYDYHFSVYSTPSESLTDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRKNPTPFEKLDF 608
              YH+SVYSTPSESLTDRFCR+D EKFG V DITDK+YYTNSFHYDVRKNPTPFEKL+F
Sbjct: 546 ASGYHYSVYSTPSESLTDRFCRMDKEKFGSVADITDKDYYTNSFHYDVRKNPTPFEKLEF 605

Query: 609 EKIYPETGASGGFIHYCEYPVLQQNPKALEAVWDYAYDRVGYLGTNTPIDKCYQCQFEGD 668
           EK YP   A+GGFIHYCEYPVLQQNPKALEAVWD+AYDR+GYLGTN PID CY C FEGD
Sbjct: 606 EKDYP-VYANGGFIHYCEYPVLQQNPKALEAVWDFAYDRIGYLGTNAPIDHCYACGFEGD 664

Query: 669 FTPTDRGFTCPNCGNSDPKTVDVVKRTCGYLGNPQARPMVNGRHKEISARVKHMNGS-SI 727
           FTPT+RGF CP CGN DPKT DVVKRTCGYLGNPQARPMV+GRHKEIS+RVKHMNGS
Sbjct: 665 FTPTERGFKCPQCGNDDPKTCDVVKRTCGYLGNPQARPMVHGRHKEISSRVKHMNGSVGA 724

Query: 728 KNQGN 732
            N GN
Sbjct: 725 LNDGN 729
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3379> which encodes the amino acid sequence <SEQ ID 3380>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2975(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 641/731 (87%), Positives = 680/731 (92%)

Query: 1    MMVLERERFMTESDIKVIKRDGRLVSFDKYKIYTALLKASNKVIKMSPLVEAKLEMIADH 60
            M+ LE ++    + DIKVIKRDGRLV+FD  KIY+ALLKAS KV +MSPLVEAKLE I+D
Sbjct: 1    MVSLEEDKVTVQPDIKVIKRDGRLVNFDSTKIYSALLKASMKVTRMSPLVEAKLEAISDR 60

Query: 61   VIAEIYNRFKDNIKIYEIQNIVEHKLLEANEYAIAQEYINYRTQRDFERSQATDINFSIG 120
            +IAEI  RF  NIKIYEIQNIVEHKLL ANEYAIA+EYINYRTQRDF RSQATDINFSI
Sbjct: 61   IIAEIIERFPTNIKIYEIQNIVEHKLLAANEYAIAKEYINYRTQRDFARSQATDINFSID 120

Query: 121  KLINKDQTVVNENANKDSDVFNTQRDLTAGIVGKSIGLKMLPSHVANAHQKGDIHYHDLD 180
            KLINKDQTVVNENANKDSDVFNTQRDLTAGIVGKSIGLKMLPSHVANAHQKGDIHYHDLD
Sbjct: 121  KLINKDQTVVNENANKDSDVFNTQRDLTAGIVGKSIGLKMLPSHVANAHQKGDIHYHDLD 180

Query: 181  YSPYTPMTNCCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTA 240
            YSPYTPMTNCCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTA
Sbjct: 181  YSPYTPMTNCCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTA 240

Query: 241  DRIDEFLAPYAQLNYQKHLKDAKEWVIEDKQEDYARAKTQKDIYDAMQSLEYEINTLFTS 300
            DRIDEFLAPYA+LN++KH+ DAK+W++E K+E YA  KTQKDIYDAMQSLEYEINTLFTS
Sbjct: 241  DRIDEFLAPYAELNFKKHMADAKKWIVETKRESYAFEKTQKDIYDAMQSLEYEINTLFTS 300

Query: 301  NGQTPFTSLGFGLGTNWFEREIQKAILKIRIQGLGSEHRTAIFPKLIFTLKKGLNLEEDS 360
```

```
             NGQTPFTSLGFGLGT+WFEREIQKAIL IRI GLGSEHRTAIFPKLIFT+K+GLNLE DS
Sbjct: 301 NGQTPFTSLGFGLGTSWFEREIQKAILTIRINGLGSEHRTAIFPKLIFTVKRGLNLEPDS 360

Query: 361 PNYDIKQLALECATKRMYPDVLSYDKIIDLTGSFKAPMGCRSFLQGWRDANGQDVTSGRM 420
            PNYDIK LALECATKRMYPD+LSYDKIIDLTGSFK+PMGCRSFLQGW+D NGQDVTSGRM
Sbjct: 361 PNYDIKTLALECATKRMYPDMLSYDKIIDLTGSFKSPMGCRSFLQGWKDENGQDVTSGRM 420

Query: 421 NLGVVTVNLPRVAMESNGDMDKFWEIFNERMSIARDALVYRVERVKEAIPANAPILYQYG 480
            NLGVVT+NLPR+AMESNGDMDKFWE+FNERM I++DAL+YRVERV EA PANAPILYQYG
Sbjct: 421 NLGVVTLNLPRIAMESNGDMDKFWELFNERMLISKDALIYRVERVTEAKPANAPILYQYG 480

Query: 481 AFGERLGKYDNVDRLFNHRRATVSLGYIGLYEVASVFYGGDWEDNHQAKAFTVDIVRKMK 540
            AFG+RL K  NV+ LF +RRATVSLGYIGLYEVASVFYGG WE N  AKAFT+ IV+ MK
Sbjct: 481 AFGKRLEKTGNVNDLFKNRRATVSLGYIGLYEVASVFYGGQWEGNPDAKAFTLSIVKAMK 540

Query: 541 QLCADWSDEYDYHFSVYSTPSESLTDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRKNP 600
            Q C DWSDEY YHFSVYSTPSESLTDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRK+P
Sbjct: 541 QACEDWSDEYGYHFSVYSTPSESLTDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRKSP 600

Query: 601 TPFEKLDFEKIYPETGASGGFIHYCEYPVLQQNPKALEAVWDYAYDRVGYLGTNTPIDKC 660
            TPFEKLDFEK YPE GASGGFIHYCEYPVLQQNPKALEAVWDYAYDRVGYLGTNTPIDKC
Sbjct: 601 TPFEKLDFEKDYPEAGASGGFIHYCEYPVLQQNPKALEAVWDYAYDRVGYLGTNTPIDKC 660

Query: 661 YQCQFEGDFTPTDRGFTCPNCGNSDPKTVDVVKRTCGYLGNPQARPMVNGRHKEISARVK 720
            Y CQFEGDFTPT+RGFTCPNCGN+DPKTVDVVKRTCGYLGNPQARPMVNGRHKEISARVK
Sbjct: 661 YNCQFEGDFTPTERGFTCPNCGNNDPKTVDVVKRTCGYLGNPQARPMVNGRHKEISARVK 720

Query: 721 HMNGSSIKNQG 731
            HMNGS+IK  G
Sbjct: 721 HMNGSTIKYPG 731
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1094

A DNA sequence (GBSx1169) was identified in *S.agalactiae* <SEQ ID 3381> which encodes the amino acid sequence <SEQ ID 3382>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5372(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3383> which encodes the amino acid sequence <SEQ ID 3384>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6084(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 28/47 (59%), Positives = 40/47 (84%), Gaps = 1/47 (2%)

Query: 1  MGKYQLDYKGQAQVQKFHEKHSTGENANQKSRLKDLRKQFLEKAKKK 47
```

```
           MGKYQLDYKG  QV++FHEKHS  +  ++KSR+++L+ +FLEK+KK+
Sbjct: 1   MGKYQLDYKGMQQVERFHEKHSK-KKTDKKSRVQELKARFLEKSKKQ 46
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1095

A DNA sequence (GBSx1170) was identified in *S.agalactiae* <SEQ ID 3385> which encodes the amino acid sequence <SEQ ID 3386>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0436(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB95794 GB:AL359949 putative oxidoreductase [Streptomyces
           coelicolor A3(2)]
 Identities = 91/299 (30%), Positives = 147/299 (48%), Gaps = 7/299 (2%)

Query: 2   LQLGIVGLGGISQKAYLPYMRQVTGVHWHLFTRQKQILEEV--NMLFGSSTAYDSLDSLA 59
           +++G +GLG I+QK YLP +  + G+  HL TR    L  V   +    +  +  LD+L
Sbjct: 1   MKVGCIGLGDIAQKGYLPVLAALPGIELHLQTRTPATLTRVADKLRIPPAQRHADLDALL 60

Query: 60  EHPLDGVFIHVATSAHFDIAKLFLKKGIPVFMDKPLTEDYTSTKALYDLAKDHKTFLMAG 119
              LD  F+H  T+AH +I    L+ G+P ++DKPL  +   ++ L  LA++  T L  G
Sbjct: 61  AQGLDAAFVHAPTAAHPEIVTRLLEAGVPTYVDKPLAYELADSERLVTLAEERGTSLAVG 120

Query: 120 FNRRFAPRIMEMKKVEDKNHIRTFKNAVNAPADFQYKLFDMFIHPLDTALFLTNNVVKRG 179
           FNRR AP   +  +   +  I   KN    P D +  + D FIH +DT  FL    V
Sbjct: 121 FNRRHAPGYAQCAE-HPRELILMQKNRTGLPEDPRTMILDDFIHVVDTLRFLVPGPVDDV 179

Query: 180 YFVTKRDGNKILQVSVTLETDSEIIEASMNLQSGSRREIIEIESPEVTYSLDDLSNLSVI 239
               + +G  +  V + L  D       MN  SGS  EI+E+   +    + +L+   VI
Sbjct: 180 TVRARTEGGLLHHVVLQLAGDGFTALGVMNRLSGSAEEILEVSGQDTKRQVVNLA--EVI 237

Query: 240 DGFDRRAI-GFGSWASTLEKRGFEPMIDAFIQAITTGVNPISPKSSLLSHFICDQINKA 297
           D   +  +   G W     +RG E  + AF+ A+ +G   +S + +L +H +C+++ +A
Sbjct: 238 DHKGQPTVRRRGDWVPVARQRGIEQAVLAFLDAVRSG-EVLSARDALATHELCERVVRA 295
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3387> which encodes the amino acid sequence <SEQ ID 3388>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF96942 GB:AE004430 oxidoreductase, Gfo/Idh/MocA family [Vibrio cholerae]
 Identities = 103/304 (33%), Positives = 158/304 (51%), Gaps = 11/304 (3%)

Query: 4   LNIGIVGLGAISQKAYLPYMRQLSDITWHLSTRNAAVRQQVGQLFGHAILYSDVKELSKT 63
           + I ++GLG I+QKAYLP + Q  DI   L TRN  V   +   +  + +D +++ +
Sbjct: 1   MKIAMIGLGDIAQKAYLPVLAQWPDIELVLCTRNPKVLGTLATRYRVSATCTDYRDVLQY 60
```

```
Query: 64  NLDGVFIHAATSAHAELASLFLNQGIPVFMDKPIADNYLMTKNLYDLAKENQTFLMAGFN 123
            +D V IHAAT  H+ LA+ FL+ GIP F+DKP+A +    +NLY+LA+++   L  GFN
Sbjct: 61  GVDAVMIHAATDVHSTLAAFFLHLGIPTFVDKPLAASAQECENLYELAEKHHQPLYVGFN 120

Query: 124 RRFTPRVKK-LSSLSTK-----RKVAVEKNDLNRPGDMTFKLFDFFIHPLDTALFLTEGT 177
           RR  P   + LS L+ +     R +  EK+    PGD+   +FD FIHPLD+     +
Sbjct: 121 RRHIPLYNQHLSELAQQECGALRSLRWEKHRHALPGDIRTFVFDDFIHPLDSVNLSRQCN 180

Query: 178 LLKGHFQYHLEAGLLSQVMVTLMTESMTTTASMNLQSGSRREVMEVQRAEETYHLENLDE 237
           L   H  YH+  GLL+++ V   T      ASMN Q G   E +        Y  ++  +
Sbjct: 181 LDDLHLTYHMSEGLLARLDVQWQTGDTLLHASMNRQFGITTEHVTASYDNVAYLFDSFTQ 240

Query: 238 LSIYKGTEKRVLGFASWDTTLHKRGFETMIDAFLEAISTGVNPVS-PESSLLSHW----I 292
             +++  ++  +    W   L  +GF+ M+  +L+  + G  P    E +L SH     I
Sbjct: 241 GKMWRDNQESRVALKDWTPMLASKGFDAMVQDWLQVAAAGKLPTHIIERNLASHQLAEAI 300

Query: 293 CQQI 296
           CQQI
Sbjct: 301 CQQI 304
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 168/308 (54%), Positives = 223/308 (71%)

Query: 1   MLQLGIVGLGGISQKAYLPYMRQVTGVHWHLFTRQKQILEEVNMLFGSSTAYDSLDSLAE 60
           ML +GIVGLG ISQKAYLPYMRQ++ + WHL TR   + ++V  LFG +  Y  +  L++
Sbjct: 3   MLNIGIVGLGAISQKAYLPYMRQLSDITWHLSTRNAAVRQQVGQLFGHAILYSDVKELSK 62

Query: 61  HPLDGVFIHVATSAHFDIAKLFLKKGIPVFMDKPLTEDYTSTKALYDLAKDHKTFLMAGF 120
             LDGVFIH ATSAH ++A LFL +GIPVFMDKP+ ++Y  TK LYDLAK+++TFLMAGF
Sbjct: 63  TNLDGVFIHAATSAHAELASLFLNQGIPVFMDKPIADNYLMTKNLYDLAKENQTFLMAGF 122

Query: 121 NRRFAPRIMEMKKVEDKNHIRTFKNAVNAPADFQYKLFDMFIHPLDTALFLTNNVVKRGY 180
           NRRF PR+ ++  +  K  +   KN +N P D  +KLFD FIHPLDTALFLT   + +G+
Sbjct: 123 NRRFTPRVKKLSSLSTKRKVAVEKNDLNRPGDMTFKLFDFFIHPLDTALFLTEGTLLKGH 182

Query: 181 FVTKRDGNKILQVSVTLETDSEIIEASMNLQSGSRREIIEIESPEVTYSLDDLSNLSVID 240
           F    +   + QV VTL T+S    ASMNLQSGSRRE++E++  E TY L++L  LS+
Sbjct: 183 FQYHLEAGLLSQVMVTLMTESMTTTASMNLQSGSRREVMEVQRAEETYHLENLDELSIYK 242

Query: 241 GFDRRAIGFGSWASTLEKRGFEPMIDAFIQAITTGVNPISPKSSLLSHFICDQINKANAP 300
           G ++R +GF SW +TL KRGFE MIDAF++AI+TGVNP+SP+SSLLSH+IC QI  +
Sbjct: 243 GTEKRVLGFASWDTTLHKRGFETMIDAFLEAISTGVNPVSPESSLLSHWICQQIADSQLS 302

Query: 301 FGMLNLKI 308
           +G L +++
Sbjct: 303 YGELTVEL 310
```

SEQ ID 3386 (GBS309) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 10; MW 63kDa).

GBS309-GST was purified as shown in Figure 212, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1096

A DNA sequence (GBSx1171) was identified in *S.agalactiae* <SEQ ID 3389> which encodes the amino acid sequence <SEQ ID 3390>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
              bacterial cytoplasm --- Certainty=0.2983(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04222 GB:AP001508 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 52/129 (40%), Positives = 70/129 (53%), Gaps = 5/129 (3%)

Query: 39  FEDWLDHNLNMELGVGVPDNFVPYIQFVSFDNDNNAIGFLNLRLRLNDTLLEKGGHIGYS 98
            FE  L    + + GV +P N V    +         IG +N+R  LND L  +GGHIGY
Sbjct: 43  FEHLLKTLKDYQHGVNLPANRVANTTYWLVHEQKRLIGAINIRHTLNDWLHHRGGHIGYG 102

Query: 99  IRPRQRGKGYAKEQLKLGIEQAHLKNINEILVTCHVDNDASKSVILANGGVLEDCLHQ-- 156
            IRP +RGKGYA   LKLG+E+A    + ++L+TC  +N  S   I  NGGVL+  +
Sbjct: 103 IRPSERGKGYATLMLKLGLEKAAALGLEKVLITCDKENLPSARTIQRNGGVLDSEVVDER 162

Query: 157 ---TERYWI 162
                +RYWI
Sbjct: 163 GIAIQRYWI 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3391> which encodes the amino acid sequence <SEQ ID 3392>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2195(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 90/164 (54%), Positives = 115/164 (69%), Gaps = 4/164 (2%)

Query: 1   MKLRRPVLEDKEEILAMYKEFQKESSSVDG--GFYEPTMHFEDWLDHNLNMELGVGVPDN 58
           M++RRP L+DK+ +L+M  EF ++ S+ DG   F     ++E WL+ +L  E+G+
Sbjct: 1   MEIRRPTLKDKDAVLSMINEFLEQKSATDGLWHFNVNDFNYETWLEDSLRQEMGLS--SQ 58

Query: 59  FVPYIQFVSFDNDNNAIGFLNLRLRLNDTLLEKGGHIGYSIRPRQRGKGYAKEQLKLGIE 118
             VP IQ+V+FD  + AIGFLNLRLRLN+ LLEKGGHIGYS+RP QRGKGYAKE LK  +
Sbjct: 59  GVPAIQYVAFDERSQAIGFLNLRLRLNERLLEKGGHIGYSVRPSQRGKGYAKEMLKQAVS 118

Query: 119 QAHLKNINEILVTCHVDNDASKSVILANGGVLEDCLHQTERYWI 162
             A  KNI  ILVTC   N AS++VI+AN G+LED    TERYWI
Sbjct: 119 YAISKNITTILVTCDETNVASRAVIVANVGILEDSRGGTERYWI 162
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1097

A DNA sequence (GBSx1172) was identified in *S.agalactiae* <SEQ ID 3393> which encodes the amino acid sequence <SEQ ID 3394>. This protein is predicted to be anaerobic ribonucleotide reductase activator protein (nrdG). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD00216 GB:U73336 anaerobic ribonucleotide reductase activator
           protein [Lactococcus lactis subsp. cremoris]
 Identities = 152/198 (76%), Positives = 176/198 (88%)

Query: 8    NTPKPGEWKSEELSHGHIIDYKAFNFVDGEGVRNSLYVAGCMFHCKGCYNTATWSFRAGI 67
            N PKPGEW+++ELS  +I DYK FNFVDGEGVR SLYV+GCMFHC+GCYN ATWSFR G
Sbjct: 2    NNPKPGEWRADELSQNYIADYKPFNFVDGEGVRCSLYVSGCMFHCEGCYNQATWSFRYGR 61

Query: 68   PYTKELEDQIMTDLEQPYVQGLTLLGGEPFLNTGILLPLLQRIRRELPEKDIWSWTGYTW 127
            PYTKELED+IM DL +PYVQGLTLLGGEPFLNT  L+PLL+RIRRELP+KDIWSWTGYTW
Sbjct: 62   PYTKELEDKIMADLAEPYVQGLTLLGGEPFLNTTFLIPLLKRIRRELPDKDIWSWTGYTW 121

Query: 128  EEMMLETQDKLEMLSLIDILVDGRFDQSKRNLMLQFRGSSNQRIIDVQKSLKEGEVVIWE 187
            EEMMLET DKLEML L+D+LVDGRF+ SK+NLMLQFRGSSNQRIIDV KS  +G+VVIWE
Sbjct: 122  EEMMLETDDKLEMLDLLDVLVDGRFELSKKNLMLQFRGSSNQRIIDVPKSRSKGQVVIWE 181

Query: 188  GLNDGDNSYEQVKRDDLL 205
             LNDG+N++EQ+ ++ L+
Sbjct: 182  KLNDGENNFEQIHKEKLI 199
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3395> which encodes the amino acid sequence <SEQ ID 3396>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4111(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 167/202 (82%), Positives = 186/202 (91%)

Query: 4    EASWNTPKPGEWKSEELSHGHIIDYKAFNFVDGEGVRNSLYVAGCMFHCKGCYNTATWSF 63
            E  WN PKP EW++EELS G IIDYKAFNFVDGEGVRNSLYV+GC+FHCKGCYN ATWSF
Sbjct: 4    EKCWNNPKPKEWQAEELSQGRIIDYKAFNFVDGEGVRNSLYVSGCLFHCKGCYNAATWSF 63

Query: 64   RAGIPYTKELEDQIMTDLEQPYVQGLTLLGGEPFLNTGILLPLLQRIRRELPEKDIWSWT 123
            +AG+PYT+ELE+QIMTDL QPYVQGLTLLGGEPFLNTGIL+PL++RIRRELPEKDIWSWT
Sbjct: 64   KAGMPYTQELEEQIMTDLAQPYVQGLTLLGGEPFLNTGILIPLIKRIRRELPEKDIWSWT 123

Query: 124  GYTWEEMMLETQDKLEMLSLIDILVDGRFDQSKRNLMLQFRGSSNQRIIDVQKSLKEGEV 183
            GYTWEEMMLET DKLEMLSLIDILVDGRFD +K+NLMLQFRGSSNQRIIDVQKSL   EV
Sbjct: 124  GYTWEEMMLETPDKLEMLSLIDILVDGRFDITKKNLMLQFRGSSNQRIIDVQKSLAAKEV 183

Query: 184  VIWEGLNDGDNSYEQVKRDDLL 205
            +IW+ LNDGD ++EQ+ R+DLL
Sbjct: 184  IIWDKLNDGDQTFEQISREDLL 205
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1098

A DNA sequence (GBSx1173) was identified in *S.agalactiae* <SEQ ID 3397> which encodes the amino acid sequence <SEQ ID 3398>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
```

```
INTEGRAL    Likelihood = -3.03   Transmembrane  102 - 118 ( 101 - 119)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD24446 GB:AF118389 unknown [Streptococcus suis]
 Identities = 97/240 (40%), Positives = 151/240 (62%), Gaps = 1/240 (0%)

Query: 2   IKILIPTAKEMKV-CQNIAWPKLSAQTKIIIDYFSTLTVSDLEDIYRINTSAARCEAQRW 60
           +KI+IP AKE+    +N ++  LS ++K ++D  S   V  +   Y++N + A  EA RW
Sbjct: 1   MKIIIPNAKEVNTNLENASFYLLSDRSKPVLDAISQFDVKKMAAFYKLNEAKAELEADRW 60

Query: 61  QDFKAKQLTLNPAIKLFNGLMYRNIKRHNLSTSEAQFMENSVFITSALYGIIPAMTLISP 120
              +  Q    PA +L++GLMYR + R  + + E  ++ + V + +ALYG+I     ISP
Sbjct: 61  YRIRTGQAKTYPAWQLYDGLMYRYMDRRGIDSKEENYLRDHVRVATALYGLIHPFEFISP 120

Query: 121 HRLDFNTKIKINNNSLKVFWRENYDTFMQSDDIMVSLLSNEFETVFSPKERQKLIHLNFI 180
           HRLDF   +KI N SLK +WR  YD  +  D++++SL S+EFE VFSP+ +++L+ + F+
Sbjct: 121 HRLDFQGSLKIGNQSLKQYWRPYYDQEVGDDELILSLASSEFEQVFSPQIQKRLVKILFM 180

Query: 181 EDRDGQLKTHSTISKKARGKCLTAMMENNCQTLEHLKQLRFDGFCYDNELSDSKQLTFVK 240
           E++ GQLK HSTISKK RG+ L+ + +NN Q L  ++  + DGF Y    S + QLTF++
Sbjct: 181 EEKAGQLKVHSTISKKGRGRLLSWLAKNNIQELSDIQDFKVDGFEYCTSESTANQLTFIR 240
```

A related GBS nucleic acid sequence <SEQ ID 10941> which encodes amino acid sequence <SEQ ID 10942> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3399> which encodes the amino acid sequence <SEQ ID 3400>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3759(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/242 (47%), Positives = 155/242 (63%)

Query: 1   MIKILIPTAKEMKVCQNIAWPKLSAQTKIIIDYFSTLTVSDLEDIYRINTSAARCEAQRW 60
           M+  LIPTAKEM + +      L   ++ I+   + +T  DL   YRI   +A+ E QRW
Sbjct: 1   MLTFLIPTAKEMTIPKESHPHLLPQDSQAILKIMAAMTTEDLAKSYRIKEESAKKEQQRW 60

Query: 61  QDFKAKQLTLNPAIKLFNGLMYRNIKRHNLSTSEAQFMENSVFITSALYGIIPAMTLISP 120
           QD  ++Q    PA +LFNGLMYR+IKR  L+T E  ++   V+ITS+ YGIIPA   I+
Sbjct: 61  QDMASQQSLAYPAYQLFNGLMYRHIKRDKLTTQEQAYLTQQVYITSSFYGIIPANHPIAE 120

Query: 121 HRLDFNTKIKINNNSLKVFWRENYDTFMQSDDIMVSLLSNEFETVFSPKERQKLIHLNFI 180
           HR DF+T+IKI   SLK +WR  Y+ F +    ++SLLS+EF+ VFS   +Q I    F+
Sbjct: 121 HRHDFHTRIKIEGQSLKSYWRPCYNQFAKEHPQVISLLSSEFDDVFSKDCKQLWISPKFM 180

Query: 181 EDRDGQLKTHSTISKKARGKCLTAMMENNCQTLEHLKQLRFDGFCYDNELSDSKQLTFVKKQ 242
            +++GQ KTHSTISKKARG  LTA MENNCQT++ LK L F GF Y  +LS   +  ++KK+
Sbjct: 181 AEKEGQFKTHSTISKKARGAFLTACMENNCQTVDSLKSLVFAGFYYHPDLSTDHEFVYIKKK 242
```

SEQ ID 3398 (GBS428) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 6; MW 30.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 4; MW 55kDa).

GBS428-GST was purified as shown in Figure 220, lane 6-7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1099

A DNA sequence (GBSx1174) was identified in *S.agalactiae* <SEQ ID 3401> which encodes the amino acid sequence <SEQ ID 3402>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.59    Transmembrane     3 -   19 (    3 -   19)

----- Final Results -----
               bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10251> which encodes amino acid sequence <SEQ ID 10252> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07024 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 86/275 (31%), Positives = 143/275 (51%), Gaps = 6/275 (2%)

Query: 17  MSYPYKANHSIESITLKVNDLENLVNFYSDIIGLTVIDKSSTRALLGVNQKIPLIILEKT 76
           M +  + N  ++ + +KV+DL   + FY +IIG  V+++S   A L  N + PL+++E+
Sbjct: 1   MEFHRQPNTFVDLVNIKVSDLSRALTFYQEIIGFQVLERSERSATLTANGRTPLLVIEQP 60

Query: 77  E---LEKHSTYGLYHTAILVPDEYHLSLALNHLLSQHIPLEGGADHGYSNAIYLSDPEGN 133
           +    ++  T GLYH A+L+P    L   LNHLL    PL+G +DH  S AIY +DP+GN
Sbjct: 61  DPVIAKQPRTTGLYHFALLLPSRADLGRFLNHLLQSGYPLQGASDHLVSEAIYFADPDGN 120

Query: 134 GIEIYNDKDISMWDIRESGQIIGITERLDIDNLLDSLVNVPNNYKLSEKTSIGHIHLSVK 193
           G+E+Y D+  S WD   +G++   TE +  +NLL    + P    L  +T +GHIHL V
Sbjct: 121 GVEVYADRPSSSWD-WSNGEVKMSTEPIHAENLLAEGKDEPWT-ALPPETILGHIHLHVA 178

Query: 194 DAKISSKLYQNVFGLDEKFAIPT-ASWIASGNYHHHLAFNNWAGPNLSKNQEDRPGISLL 252
           +   +   Y    G +    +   A +I++GNYHHH+  N W G       E   G+
Sbjct: 179 NLFEAETFYIEGLGFNVVARLGNQALFISTGNYHHHIGLNTWNGVGAPTPPEHSVGLKWF 238

Query: 253 TIAYNDDNLFRDSLKKAQLYQLTFLEKQDHYYIIE 287
           ++ Y  + +   ++ + +        K    ++I+
Sbjct: 239 SLTYPSEEVRAKTVNRLETIGFQVERKHGEEWVID 273
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3403> which encodes the amino acid sequence <SEQ ID 3404>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0936(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 143/282 (50%), Positives = 194/282 (68%)

Query: 17  MSYPYKANHSIESITLKVNDLENLVNFYSDIIGLTVIDKSSTRALLGVNQKIPLIILEKT 76
           M YPY +  S+ +++L V DL  +  FY+ IIGL V+ + +T   L  + K  ++ L +T
Sbjct: 1   MIYPYNSTISLGTVSLNVTDLAKMTTFYTSIIGLQVLSQDTTSRQLTTDGKTVILELRQT 60

Query: 77  ELEKHSTYGLYHTAILVPDEYHLSLALNHLLSQHIPLEGGADHGYSNAIYLSDPEGNGIE 136
            L     YGLYHTA LVPD + L L LNH L++ I LEG ADHG+S AIYLSDPEGNGIE
Sbjct: 61  PLPGDKAYGLYHTAFLVPDRHSLGLVLNHFLTRSISLEGAADHGHSEAIYLSDPEGNGIE 120

Query: 137 IYNDKDISMWDIRESGQIIGITERLDIDNLLDSLVNVPNNYKLSEKTSIGHIHLSVKDAK 196
           IY+DK +  WDIR++GQIIG+TE  D  ++L+ L ++P ++ L++ T I H+HLSVK+A
Sbjct: 121 IYHDKAVEHWDIRDNGQIIGVTEPTDTKSILEQLTDIPKHFLLAQDTRIRHVHLSVKNAL 180

Query: 197 ISSKLYQNVFGLDEKFAIPTASWIASGNYHHHLAFNNWAGPNLSKNQEDRPGISLLTIAY 256
            SS LYQ VF L +K  IP+ASWIASGNY+HHLAFN+W+ P L K+QE  PG++ LTI
Sbjct: 181 ASSLLYQKVFDLGDKMTIPSASWIASGNYYHHLAFNHWSAPYLKKHQEGAPGLAFLTIHI 240

Query: 257 NDDNLFRDSLKKAQLYQLTFLEKQDHYYIIEDFDGIRIKVVL 298
               LF  +LKKA+L+ L  L++    +  ED +GIR+ V+L
Sbjct: 241 ETPLLFSATLKKARLHGLAILQEDSSSFTTEDEEGIRVNVIL 282
```

SEQ ID 3402 (GBS429) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 7; MW 34.2kDa).

GBS429-His was purified as shown in Figure 214, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1100

A DNA sequence (GBSx1175) was identified in *S.agalactiae* <SEQ ID 3405> which encodes the amino acid sequence <SEQ ID 3406>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2362(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10249> which encodes amino acid sequence <SEQ ID 10250> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC21682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]
 Identities = 89/261 (34%), Positives = 151/261 (57%), Gaps = 4/261 (1%)

Query: 10  MVRLIFSDIDGTLINSNFKVTPKTRQGIKQIVAQGATFVPISARMPEAITPIMEQIGIDS 69
           M + +FSD +GTL+ S   ++P+T   IK++ A G  FVPISAR P  I P  +Q+  ++
Sbjct: 2   MYKAVFSDFNGTLLTSQHTISPRTVVVIKRLTANGIPFVPISARSPLGILPYWKQLETNN 61

Query: 70  YIISYNGALIQDMQQKTIASHTMDGQVALQVCSYVSKHYSKIAWNVYRYHEWYSCDKENE 129
            +++++GALI +   + I S  ++ +  L++ + +++H   +  N Y  ++ ++ D EN+
Sbjct: 62  VLVAFSGALILNQNLEPIYSVQIEPKDILEINTVLAEH-PLLGVNYYTNNDCHARDVENK 120

Query: 130 WVQKEEEIVGLQSKEMSLMELEKQDRIHKLLLMGEPSLMGELENTLKAQYPHLSIAQSAP 189
           WV  E  +  ++      +        HK+ ++GE   + E+E  LK ++PHLSI +S
Sbjct: 121 WVIYERSVTKIEIHPFDEVATRSP---HKIQIIGEAEEIIEIEVLLKEKFPHLSICRSHA 177
```

```
Query: 190 YFIEIMAPGIEKGKSAKTLADYLDISLADSIAFGDNYNDLNLLEIVGKGFVMGNAPKDLQ 249
            F+E+M     KG + + L DY   +   + IAFGDN+NDL++LE VG G  MGNAP +++
Sbjct: 178 NFLEVMHKSATKGSAVRFLEDYFGVQTNEVIAFGDNFNDLDMLEHVGLGVAMGNAPNEIK 237

Query: 250 ERIGNVTQDNDNDGIYYALVE 270
           +    VT  N+ DG+   L E
Sbjct: 238 QAANVVTATNNEDGLALILEE 258
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1101

A DNA sequence (GBSx1176) was identified in *S.agalactiae* <SEQ ID 3409> which encodes the amino acid sequence <SEQ ID 3410>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG07223 GB:AE004801 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 103/283 (36%), Positives = 165/283 (57%), Gaps = 1/283 (0%)

Query: 33  KHIGILQYVEHPSLTATRKGFIKELAKEGYKDGKNIKIEYKNAQGDQSNIQSISEKLIKD 92
           K + +   VEHP+L A R G  + L + GY+DGKN+K +Y++AQG+      I+ K I D
Sbjct: 31  KSVAVTAIVEHPALDAARDGVKEALQEAGYEDGKNLKWQYQSAQGNTGTAAQIARKFIGD 90

Query: 93  NK-LVLGIATPAAQSLTTVSTETPILFTAVTDPVSAELVKSMKKPEGLATGTSDMSPIKK 151
              +++GIATP+AQ+L   +   PI+F+ VTDPV A L  S +      TG SDM  + K
Sbjct: 91  KPDVIVGIATPSAQALVAATKSIPIVFSTVTDPVGAHLTPSWEASGTNVTGVSDMLALDK 150

Query: 152 QVSLLRKVMPKVKRVGIMYTTSERNSEVQVKQAKKIFQEAGIKTSVKGISSTNDVQDTAK 211
           Q+ L++KV+P  KR+G++Y   E NS V VK+ K++  + G+         + DV   A+
Sbjct: 151 QIELIKKVVPGAKRIGMVYNPGEANSVVVVKELKELLPKMGLSLVEASAPRSVDVSSAAR 210

Query: 212 SLMSKTEVIFVPTDNIIASSVTLLGNLSKELKVPVVGGSADMVPSGLLFSYGADYEALGR 271
           SL+ K + I+  TDN + S+   L  +  + K+P++    D V  G + + G +Y+ +G+
Sbjct: 211 SLVGKVDAIYTNTDNNVVSAYEALVKVGNDAKIPLIASDTDSVKRGAIAALGINYKEMGK 270

Query: 272 QTARQAVKILKGKDVAKVPSEYPQNLKVVVNEDMAKELGIDVS 314
           QT R  V+ILKG+   ++  E   NL++ VN   A++ G+ +S
Sbjct: 271 QTGRMVVRILKGEKPGEIKPETSDNLQLFVNPGAAQKQGVTLS 313
```

There is also homology to SEQ ID 2712.

SEQ ID 3410 (GBS188) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 2; MW 36.6kDa).

The GBS188-His fusion product was purified (Figure 204, lane 6) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 247), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1102

A DNA sequence (GBSx1177) was identified in *S.agalactiae* <SEQ ID 3411> which encodes the amino acid sequence <SEQ ID 3412>. This protein is predicted to be probable permease of ABC transporter (rbsC). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-16.13   Transmembrane  132 - 148 ( 124 - 160)
   INTEGRAL    Likelihood = -6.42   Transmembrane  241 - 257 ( 238 - 258)
   INTEGRAL    Likelihood = -6.32   Transmembrane  264 - 280 ( 260 - 284)
   INTEGRAL    Likelihood = -6.00   Transmembrane  213 - 229 ( 207 - 235)
   INTEGRAL    Likelihood = -4.67   Transmembrane   58 -  74 (  57 -  75)
   INTEGRAL    Likelihood = -1.38   Transmembrane   36 -  52 (  36 -  52)
   INTEGRAL    Likelihood = -0.85   Transmembrane   90 - 106 (  87 - 106)

----- Final Results -----
               bacterial membrane --- Certainty=0.7453(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG07224 GB:AE004801 probable permease of ABC transporter
           [Pseudomonas aeruginosa]
 Identities = 114/285 (40%), Positives = 175/285 (61%), Gaps = 3/285 (1%)

Query: 5   ILSGISQGLLWSIMAIGVFITFRILDIADLSAEGAFPMGAAVCALCIVNDINPIVATIAG 64
           +   +  GL++S++A+GVFI+FR+L   DL+ +G+FP+G AVCA  I     +P  AT+A
Sbjct: 6   LFGALEIGLIFSLVALGVFISFRLLRFPDLTVDGSFPLGGAVCATLIALGWDPYSATLAA 65

Query: 65  MLGGMLAGLVSGFLHTKMKIPALLTGIITLTGLYSINLLVLGRSNVSFALKNTLVTMVTR 124
              G LAGL +G L+ K+KI  LL  I+ +  LYSINL ++G+ NV    + TL T++
Sbjct: 66  TAAGALAGLATGLLNVKLKIMDLLASILMMIALYSINLRIMGKPNVPLIAEPTLFTLLQP 125

Query: 125 LGLNKLSAVLLIGIVCVGLVILILYLFLNTQLGLALRATGDNEAMGQANSIKVDRMKMLG 184
             L+      L+ +  V    L+L  F  TQ GLA+RATG N  M +A  +    M +LG
Sbjct: 126 EWLSDYVFRPLLLVFIVIAAKLLLDWFFTTQKGLAIRATGSNPRMARAQGVNTGGMILLG 185

Query: 185 YMIGNGLIALSGALLAQNNGYADLNMGVGTIVIGLASIILAEVMIKYLPLGKRLWSIVLG 244
             I N L+AL+GAL AQ  G AD++MG+GTIVIGLA++I+ E ++    L     +++LG
Sbjct: 186 MAISNALVALAGALFAQTQGGADISMGIGTIVIGLAAVIVGESILPSRRLILATLAVILG 245

Query: 245 SVLYRMIIVFILTTD---IDAQMIKLVSAILLALILYVPELRAKL 286
           +++YR  I   L +D   + AQ + LV+A+L+ + L +P ++ +L
Sbjct: 246 AIVYRFFIALALNSDFIGLQAQDLNLVTAVLVTVALVIPMMKKRL 290
```

There is also homology to SEQ ID 2716.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1103

A DNA sequence (GBSx1178) was identified in *S.agalactiae* <SEQ ID 3413> which encodes the amino acid sequence <SEQ ID 3414>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3798(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF86640 GB:AF162694 ABC transporter [Enterococcus gallinarum]
 Identities = 171/264 (64%), Positives = 213/264 (79%), Gaps = 1/264 (0%)

Query: 3   LLELVNLHKTFEKGTVNENHVLRGLDLTIEDGDFISVIGGNGAGKSTLLNCIAGLIPIDQ 62
            +L + +LH+TFEKGT+NENHVLRG+DLT+  GDFI++IGGNGAGKSTLLN IAG IP +Q
Sbjct: 5   VLTISDLHQTFEKGTINENHVLRGIDLTMNSGDFITIIGGNGAGKSTLLNSIAGTIPTEQ 64

Query: 63  GAITLDNQSITKDSVEKRSKDISRVFQDPRMGTATNLTIEENMAIAHKRGNKRHIFRQSV 122
            G I L ++ IT+ SV +RSK+ISRVFQDPRMGTA  LT+EEN+A+A+KRG  R  F   V
Sbjct: 65  GKIVLGDKEITRHSVTRRSKEISRVFQDPRMGTAVRLTVEENLALAYKRGQVRG-FSSGV 123

Query: 123 TDDDRQLFKKSLSQLGLGLENRMKTDAAFLSGGQRQALTLAMATLVRPKLLLLDEHTAAL 182
                R  FK+ L++L LGLENR+ T+   LSGGQRQA+TL MATL +PKL+LLDEHTAAL
Sbjct: 124 KGKHRAFFKEKLARLNLGLENRLTTEIGLLSGGQRQAITLLMATLQQPKLILLDEHTAAL 183

Query: 183 DPKTSDMVMELTQKVIEEQRLTALMITHNMEHAIAYGNRLVMLYHGKIVVDVKGEAKRNL 242
            DPKTS  VM LT ++I+EQ+LTA M+TH+ME AI YGNRL+ML+ GKIVVD+ GE K++L
Sbjct: 184 DPKTSMTVMALTDQLIQEQQLTAFMVTHDMEDAIRYGNRLIMLHQGKIVVDITGEEKQSL 243

Query: 243 TVAELMELFHKNSGQQLIDDALVL 266
            TV +LM LFH+NSG +L DD L+L
Sbjct: 244 TVPDLMALFHQNSGTELKDDQLLL 267
```

There is also homology to SEQ ID 2720:

```
Identities = 116/249 (46%), Positives = 166/249 (66%), Gaps = 1/249 (0%)

Query: 3   LLELVNLHKTFEKGTVNENHVLRGLDLTIEDGDFISVIGGNGAGKSTLLNCIAGLIPIDQ 62
            ++EL+N     + G  +   +L  + LTI + DF++++GGNGAGKSTL N IAG + + +
Sbjct: 4   IIELINATVDVDNGFEDAKTILDNVTLTIYEHDFLTILGGNGAGKSTLFNVIAGTLSLTR 63

Query: 63  GAITLDNQSITKDSVEKRSKDISRVFQDPRMGTATNLTIEENMAIAHKRGNKRHIFRQSV 122
            G I +  Q +T    EKR+  +SRVFQD +MGTA  +T+ EN+ IA +RG KR +  + +
Sbjct: 64  GQIRILGQDVTHWPAEKRALYLSRVFQDSKMGTAPRMTVAENLLIARQRGGKRSLASRKI 123

Query: 123 TDDDRQLFKKSLSQLGLGLENRMKTDAAFLSGGQRQALTLAMATLVRPKLLLLDEHTAAL 182
            T+     F+  + + G GLE  ++T A  LSGGQRQAL+L MATL +P LLLLDEHTAAL
Sbjct: 124 TEHLAS-FEDLVKRTGNGLEKHLETPAGLLSGGQRQALSLLMATLKKPALLLLDEHTAAL 182

Query: 183 DPKTSDMVMELTQKVIEEQRLTALMITHNMEHAIAYGNRLVMLYHGKIVVDVKGEAKRNL 242
            DPKTS  +M+LT + + +  LTALMITH+ME A+ YGNRL+++  G I+ D+    K  L
Sbjct: 183 DPKTSQSLMQLTDEFVTKDGLTALMITHHMEDALTYGNRLIVMKDGNIIKDLNQMEKEQL 242

Query: 243 TVAELMELF 251
            T+ +  +LF
Sbjct: 243 TITDYYQLF 251
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1104

A DNA sequence (GBSx1179) was identified in *S.agalactiae* <SEQ ID 3415> which encodes the amino acid sequence <SEQ ID 3416>. This protein is predicted to be mannose-specific phosphotransferase system component IIAB. Analysis of this protein sequence reveals the following:

```
Possible site: 54
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3527(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46485 GB:AF130465 mannose-specific phosphotransferase system
           component IIAB [Streptococcus salivarius]
 Identities = 287/336 (85%), Positives = 306/336 (90%), Gaps = 6/336 (1%)

Query: 1   MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTFMPNEGPDDLYGHFNNAIAQFDADD 60
           MGIGIIIASHGKFAEGIHQSGSMIFG+QEKVQVVTFMP+EGPDDLY HFN+AIAQFDADD
Sbjct: 1   MGIGIIIASHGKFAEGIHQSGSMIFGDQEKVQVVTFMPSEGPDDLYAHFNDAIAQFDADD 60

Query: 61  EVLVLADLWSGSPFNQASRVMGENPERKMAIITGLNLPMLIQAYTERMMDANAGVEQVAA 120
           E+LVLADLWSGSPFNQASR+ GENP+RK+AIITGLNLPMLIQAYTERMMDANA  EQVAA
Sbjct: 61  EILVLADLWSGSPFNQASRIAGENPDRKIAIITGLNLPMLIQAYTERMMDANATAEQVAA 120

Query: 121 NIIKESKEGIKALPEELNPVVEATPVAGVPADVPAEVKQSGSIPEGTVIGDGKLKINLAR 180
           NIIKE+K GIKALPEELNP  E T  A V A  P      G+IPEGTVIGDGKLKINLAR
Sbjct: 121 NIIKEAKGGIKALPEELNPAEETT-AAPVEAAAP-----QGAIPEGTVIGDGKLKINLAR 174

Query: 181 IDTRLLHGQVATAWTPASKANRIIVASDEVSKDELRKQLIKQAAPGGVKANVVPISKLIE 240
           +DTRLLHGQVAT WTPASKA+RIIVASD+V+KDELRK+LIKQAAP GVKANVVPI KLI+
Sbjct: 175 LDTRLLHGQVATNWTPASKADRIIVASDDVAKDELRKELIKQAAPNGVKANVVPIQKLID 234

Query: 241 VAKDPRFGNTRALILFETVQDALRAIEGGVEIPELNVGSMAHSTGKTMVNNVLSMDKDDV 300
            +KDPRFGNT ALILFETVQDALRAIEGGV I ELNVGSMAHSTGKTMVNNVLSMDKDDV
Sbjct: 235 ASKDPRFGNTHALILFETVQDALRAIEGGVPIKELNVGSMAHSTGKTMVNNVLSMDKDDV 294

Query: 301 AAFEKLRDLGVSFDVRKVPNDAKKNLFDLINKANVK 336
           A FEKLRDLGV FDVRKVPND+KK+LFDLI KANV+
Sbjct: 295 ACFEKLRDLGVEFDVRKVPNDSKKDLFDLIKKANVQ 330
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3417> which encodes the amino acid sequence <SEQ ID 3418>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3533(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 288/336 (85%), Positives = 308/336 (90%), Gaps = 6/336 (1%)

Query: 1   MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTFMPNEGPDDLYGHFNNAIAQFDADD 60
           MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTFMPNEGPDDLYGHFNNAI QFDADD
Sbjct: 1   MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTFMPNEGPDDLYGHFNNAIQQFDADD 60

Query: 61  EVLVLADLWSGSPFNQASRVMGENPERKMAIITGLNLPMLIQAYTERMMDANAGVEQVAA 120
           E+LVLADLWSGSPFNQASRV GENP+RKMAIITGLNLPMLIQAYTER+MDA AGVEQVAA
Sbjct: 61  EILVLADLWSGSPFNQASRVAGENPDRKMAIITGLNLPMLIQAYTERLMDAGAGVEQVAA 120

Query: 121 NIIKESKEGIKALPEELNPVVEATPVAGVPADVPAEVKQSGSIPEGTVIGDGKLKINLAR 180
           NIIKESK+GIKALPE+LNPV E      V   +       G+IP GTVIGDGKLKINLAR
Sbjct: 121 NIIKESKDGIKALPEDLNPVEETAATEKVVNAL------QGAIPAGTVIGDGKLKINLAR 174

Query: 181 IDTRLLHGQVATAWTPASKANRIIVASDEVSKDELRKQLIKQAAPGGVKANVVPISKLIE 240
           +DTRLLHGQVATAWTPASKA+RIIVASDEV++D+LRKQLIKQAAPGGVKANVVPISKLIE
Sbjct: 175 VDTRLLHGQVATAWTPASKADRIIVASDEVAQDDLRKQLIKQAAPGGVKANVVPISKLIE 234
```

```
Query: 241 VAKDPRFGNTRALILFETVQDALRAIEGGVEIPELNVGSMAHSTGKTMVNNVLSMDKDDV 300
            +KDPRFGNT ALILF+T QDALRA+EGGVEI ELNVGSMAHSTGKTMVNNVLSMDK+DV
Sbjct: 235 ASKDPRFGNTHALILFQTPQDALRAVEGGVEINELNVGSMAHSTGKTMVNNVLSMDKEDV 294

Query: 301 AAFEKLRDLGVSFDVRKVPNDAKKNLFDLINKANVK 336
           A FEKLRDLGV+FDVRKVPND+KKNLF+LI K N+K
Sbjct: 295 ATFEKLRDLGVTFDVRKVPNDSKKNLFELIQKTNIK 330
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1105

A DNA sequence (GBSx1180) was identified in *S.agalactiae* <SEQ ID 3419> which encodes the amino acid sequence <SEQ ID 3420>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3873(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06625 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 89/267 (33%), Positives = 139/267 (51%), Gaps = 3/267 (1%)

Query: 3   KKIIAVDLDGTLLHNNNTISDYTADTLRKVQAQGHKVIITTGRPYRMALAHYLRLDLKTP 62
            + +IA+DLDGTLL +N TIS  T  T++K +  GH V+I+TGRPYR ++ +Y  L L T
Sbjct: 4   RHLIALDLDGTLLTDNKTISMKTKQTIQKAREAGHIVVISTGRPYRASIQYYQELQLDTA 63

Query: 63  MINFNGALTHIPEKKWAFERSATIDKKLLLETLNLSDAIQADFIASEYRKNFYITMDNRD 122
            ++NFNGA  H P+          ++     + +   +A     I  E   ++Y+     D
Sbjct: 64  IVNFNGAFVHHPKDSSFGTYHHPLELSTARQVIETCEAFDVSNIMVEVIDDYYLRY--YD 121

Query: 123 KINPQLFGVNEITDKMALDVTKITRNPNALLMQTRHKDKYELAKELRQHFNHELEVDSWG 182
            ++  Q F   +   +    + K+  +P  +L+  +     EL   L       ++  +WG
Sbjct: 122 ELFIQTFTEGQGPVEHGNLLKKLRDDPTCVLIHPKDDHVSELRSLLDGAHAEVIDQRTWG 181

Query: 183 GPLNILEFSPKGVNKAYALKHLLKSLNLSQENLIAFGDEHNDTEMLAFAHTGYAMKNANP 242
             P N++E    G+NKA  LK +     + +E +IAFGDE ND EM+ +A  G AM NA
Sbjct: 182 APWNVIEIVKAGMNKAVGLKRIADYYQVPKERIIAFGDEDNDFEMIEYAGKGVAMANAID 241

Query: 243 TLLPYADQQIQWTNEEDGVAKTLEKLL 269
             L   A+  I  +NE+DG+A  LE+ L
Sbjct: 242 PLKALAN-DITLSNEDDGIAVYLEEAL 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3421> which encodes the amino acid sequence <SEQ ID 3422>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4380(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 188/270 (69%), Positives = 224/270 (82%)
```

```
Query: 1   MTKKIIAVDLDGTLLHNNNTISDYTADTLRKVQAQGHKVIITTGRPYRMALAHYLRLDLK 60
           MTKK+IA+DLDGTLLH++NTIS YT  T++ VQ +GH VII+TGRPYRMAL +YL+L+LK
Sbjct: 1   MTKKLIAIDLDGTLLHHDNTISTYTQKTIKAVQDKGHHVIISTGRPYRMALGYYLQLNLK 60

Query: 61  TPMINFNGALTHIPEKKWAFERSATIDKKLLLETLNLSDAIQADFIASEYRKNFYITMDN 120
           TP+I FNGALTH+PE+KWA+E + T+DK  LL  L   D  Q DFIASEYRKN YITM N
Sbjct: 61  TPIITFNGALTHMPEQKWAYEHNVTLDKGYLLRLLKYQDDFQMDFIASEYRKNVYITMTN 120

Query: 121 RDKINPQLFGVNEITDKMALDVTKITRNPNALLMQTRHKDKYELAKELRQHFNHELEVDS 180
            + I+PQLFGV+EIT  MAL++TKITRNPNALLMQT H+DKY LAK +R  F  E+E+DS
Sbjct: 121 PESIDPQLFGVDEITQDMALEITKITRNPNALLMQTHHEDKYALAKNMRACFKDEIEIDS 180

Query: 181 WGGPLNILEFSPKGVNKAYALKHLLKSLNLSQENLIAFGDEHNDTEMLAFAHTGYAMKNA 240
           WGGPLNILE S K VNKAYAL +LL   N+ +++LIAFGDEHNDTEMLAFA TGYAMKNA
Sbjct: 181 WGGPLNILEISSKNVNKAYALNYLLGIYNMDKKDLIAFGDEHNDTEMLAFAGTGYAMKNA 240

Query: 241 NPTLLPYADQQIQWTNEEDGVAKTLEKLLL 270
           +P LLPYADQQ+ ++NEEDGVAK LE+L L
Sbjct: 241 SPVLLPYADQQLNFSNEEDGVAKKLEELFL 270
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1106

A DNA sequence (GBSx1181) was identified in *S.agalactiae* <SEQ ID 3423> which encodes the amino acid sequence <SEQ ID 3424>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.38   Transmembrane   96 - 112 (  90 - 119)
   INTEGRAL    Likelihood = -6.58   Transmembrane   28 -  44 (  27 -  47)
   INTEGRAL    Likelihood = -6.26   Transmembrane  176 - 192 ( 174 - 193)
   INTEGRAL    Likelihood = -5.26   Transmembrane  127 - 143 ( 126 - 144)
   INTEGRAL    Likelihood = -1.59   Transmembrane    4 -  20 (   3 -  20)
   INTEGRAL    Likelihood = -0.22   Transmembrane   60 -  76 (  59 -  78)

----- Final Results -----
               bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1107

A DNA sequence (GBSx1182) was identified in *S.agalactiae* <SEQ ID 3425> which encodes the amino acid sequence <SEQ ID 3426>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2025(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1108

A DNA sequence (GBSx1183) was identified in *S.agalactiae* <SEQ ID 3427> which encodes the amino acid sequence <SEQ ID 3428>. This protein is predicted to be an integral membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.41   Transmembrane  180 - 196 ( 179 - 199)
   INTEGRAL    Likelihood = -5.31   Transmembrane   96 - 112 (  94 - 114)
   INTEGRAL    Likelihood = -2.18   Transmembrane  129 - 145 ( 129 - 145)
   INTEGRAL    Likelihood = -1.33   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8729> which encodes amino acid sequence <SEQ ID 8730> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      5.85
GvH: Signal Score (-7.5): -2.39
     Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 4 value:  -5.41 threshold:  0.0
   INTEGRAL    Likelihood = -5.41   Transmembrane  176 - 192 ( 175 - 195)
   INTEGRAL    Likelihood = -5.31   Transmembrane   92 - 108 (  90 - 110)
   INTEGRAL    Likelihood = -2.18   Transmembrane  129 - 145 ( 129 - 145)
   PERIPHERAL  Likelihood =  0.05      57
 modified ALOM score:   1.58

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC65028 GB:AE001188 conserved hypothetical integral membrane
           protein [Treponema pallidum]
 Identities = 54/190 (28%), Positives = 93/190 (48%), Gaps = 14/190 (7%)

Query: 14  LFFIVISFGIKYYHLQG--PNLIWNMTLALIALDFAYLTSL--FKKKILIGLFALAWFFF 69
           +F +++SFG +          L+WN+ LA I    + +  +  F  + +     L W  F
Sbjct: 3   VFCLLLSFGRRCVAADNFLSFLVWNLVLAFIPWLISAILHVRRFAVRSVQLFLMLLWLLF 62

Query: 70  YPNTFYMLTDIIHMHFVGDVLYNKTNLILYILYVSSILFGFLSGIESFSVIMRKFRISNI 129
           +PN  Y+LTDIIH+      L     +IL     + + + F+S     S++ R F I
Sbjct: 63  FPNAPYILTDIIHLGKGKSFLLYYDLIILLAYSFTGLFYAFVSLHLIESILARDFHIKRP 122

Query: 130 FLRWGIIGIVSL-VSSFGIHIGRYARLNSWDILTKPQVVINELLAVPSR-----DSFHFI 183
           F    II +  L + +FGI++GR+ R NSWDI+   + +++++     R     D++ F+
Sbjct: 123 F----IISVFELYLCAFGIYLGRFLRWNSWDIVLHGRTILSDIGIRVIRPVFYVDTWMFV 178

Query: 184 LGFTFLQVLC 193
```

```
                  F  + VLC
Sbjct: 179 FFFGTMLVLC 188
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1109

A DNA sequence (GBSx1184) was identified in *S.agalactiae* <SEQ ID 3429> which encodes the amino acid sequence <SEQ ID 3430>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.79   Transmembrane  171 - 187 ( 166 - 191)

----- Final Results -----
               bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1110

A DNA sequence (GBSx1185) was identified in *S.agalactiae* <SEQ ID 3431> which encodes the amino acid sequence <SEQ ID 3432>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.46   Transmembrane  193 - 209 ( 191 - 214)
   INTEGRAL    Likelihood =-10.30   Transmembrane   99 - 115 (  96 - 119)
   INTEGRAL    Likelihood = -8.17   Transmembrane  454 - 470 ( 451 - 472)
   INTEGRAL    Likelihood = -6.64   Transmembrane  216 - 232 ( 212 - 236)
   INTEGRAL    Likelihood = -6.37   Transmembrane   49 -  65 (  43 -  68)
   INTEGRAL    Likelihood = -4.88   Transmembrane  362 - 378 ( 357 - 383)
   INTEGRAL    Likelihood = -3.61   Transmembrane  385 - 401 ( 385 - 402)
   INTEGRAL    Likelihood = -2.76   Transmembrane  275 - 291 ( 275 - 291)
   INTEGRAL    Likelihood = -1.70   Transmembrane   18 -  34 (  18 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF95422 GB:AE004299 conserved hypothetical protein [Vibrio cholerae]
 Identities = 193/471 (40%), Positives = 286/471 (59%), Gaps = 42/471 (8%)

Query: 1    MEKFFKLKEHGTTIRTEITAGLTTFFAMSYILFVNPAILSQTGMPAQGVFLATIIGAVVA 60
             +EK FKL E+GT +RTEI AG+TTF  M+YI+FVNPAILS  GM    VF+AT + A +
Sbjct: 2    LEKLFKLSEYGTNVRTEILAGVTTFLTMAYIIFVNPAILSDAGMDRGAVFVATCLAAAIG 61

Query: 61   TSVMAFYANLPYAQAPGMGLNAFFTYTVVFALGYTWQEALAMVFICGLISLIITLTKVRK 120
               +M F AN P AQAPGMGLNAFFTY VV  +G+TWQ ALA VF  G++ ++++L K+R+
Sbjct: 62   CFIMGFIANYPIAQAPGMGLNAFFTYGVVLGMGHTWQVALAAVFCSGVLFILLSLFKIRE 121
```

```
Query: 121 MIIESIPTTLKSAITAGIGTFLAYVGIKNAGFLKFSIDPGTYDVVGKGAAKGLATITANS 180
            II SIP +L++ I+AGIG FLA++ +KNAG +    +P T  +V  GA   L  +
Sbjct: 122 WIINSIPHSLRTGISAGIGLFLAFIALKNAGIV--VDNPAT--LVSLGAITSLHAV---- 173

Query: 181 SATPGLVSFDNPAILLSLIGLSITIFFIVKGIRGGIILSILTTTLLGILMGVVKLDAINW 240
                          L+ +G  +TI  + +G++G ++++IL  T LG++ G V+   I
Sbjct: 174 ---------------LAAVGFFLTIGLVYRGVKGAVMIAILAVTALGLVFGDVQWGGIMS 218

Query: 241 EATNLSASFRDLKQVFGVALGEKGLISLFSNPSRLPSVLMAILAFSLTDIFDTIGTLIGT 300
               +++ +F    Q+   A+ E G+IS+             + AF   D+FDT GTL+G
Sbjct: 219 TPPSIAPTF---MQLDFSAVFEIGMISV-------------VFAFLFVDLFDTAGTLVGV 262

Query: 301 GEKVGILATTGDNHESKSLDKALYSDLIGTTFGAICGTSNVTTYVESAAGIGAGGRTGLT 360
              K G++   G   +   L++AL +D   T+ GA+ GTSN T+Y+ES +G+  GGRTGLT
Sbjct: 263 ATKAGLIEKDG---KIPRLNRALLADSTATSVGALLGTSNTTSYIESVSGVAVGGRTGLT 319

Query: 361 ALVVAGLFAISSFFSPLVSIVPSQATAPILVIVGIMMLSNLKDIKWDDMSEAIPAFFTSL 420
           A+VV  LF ++ FFSPL  ++P+ ATA  L  V I+M+S L  I W D++EA P   T L
Sbjct: 320 AVVVGILFLLALFFSPLAGMIPAYATAGALFYVAILMMSGLVSIDWRDLTEAAPTVVTCL 379

Query: 421 FMGFTYSITYGIAAGFLTYTLAKVIKGQAKDIHVVLWILDILFILNFISLA 471
            M  T+SI  GI+ GF+ Y   K+  G+ + + + +W++  +F++ +I  A
Sbjct: 380 MMPLTFSIAEGISLGFIAYAAIKLFSGKGRSVSLSVWVMAAIFVIKYILAA 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3433> which encodes the amino acid sequence <SEQ ID 3434>. Analysis of this protein sequence reveals the following:

```
      Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.57    Transmembrane  378 - 394 ( 370 - 419)
   INTEGRAL    Likelihood = -9.29    Transmembrane  202 - 218 ( 195 - 221)
   INTEGRAL    Likelihood = -7.64    Transmembrane   48 -  64 (  46 -  71)
   INTEGRAL    Likelihood = -7.64    Transmembrane   99 - 115 (  97 - 118)
   INTEGRAL    Likelihood = -6.90    Transmembrane  225 - 241 ( 221 - 245)
   INTEGRAL    Likelihood = -6.05    Transmembrane  468 - 484 ( 465 - 485)
   INTEGRAL    Likelihood = -4.35    Transmembrane  399 - 415 ( 395 - 419)
   INTEGRAL    Likelihood = -3.24    Transmembrane  425 - 441 ( 425 - 442)
   INTEGRAL    Likelihood = -3.08    Transmembrane   18 -  34 (  18 -  34)
   INTEGRAL    Likelihood = -2.28    Transmembrane  442 - 458 ( 442 - 460)
   INTEGRAL    Likelihood = -0.00    Transmembrane  282 - 298 ( 282 - 298)

----- Final Results -----
               bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04327 GB:AP001509 unknown conserved protein [Bacillus halodurans]
 Identities = 192/485 (39%), Positives = 276/485 (56%), Gaps = 53/485 (10%)

Query: 1   MEKFFKLSENGTTVSTEIMAGLTTFFAMSYILFVNPSILGAAGMPSNAVFLATIIAAAIS 60
           M+++F   E+GTT   E +AGLTTF +M+YILFVNP ILG AGM   AVF+AT +AAAI
Sbjct: 1   MDRYFGFKEHGTTYGRESIAGLTTFLSMAYILFVNPLILGDAGMDVQAVFMATALAAAIG 60

Query: 61  TLIMGLFANVPYALAPGMGLNAFFTYTVVFALRFSWQEALAMVFICGLFNIFITVTKFRK 120
           TLIMG+ A  P ALAPGMGLNAFF Y+VV  +   WQ AL  VF+ G+  I ITV K R+
Sbjct: 61  TLIMGILAKYPIALAPGMGLNAFFAYSVVIGMGIDWQLALFGVFVSGIIFILITVFKIRE 120

Query: 121 SIIKAIPVSLQHAIGGGIGVFVAYLGFKNANIITFSISAENIVMVNGVEPAKASAKTFAD 180
            II AIP  L++A   GIG+F+A++G KNA I+
Sbjct: 121 VIINAIPAELKNAAAAGIGLFIAFIGLKNAGIVV-------------------------- 154

Query: 181 GLLFVDANGGVVPTISSFTDSGVLLAIFGLLLTTALVIRNFRGAILIGIVATTLVGIPLG 240
                 ++     ++    +   LLA FGL++T   ++R  +G I  G++ T +VG+  G
Sbjct: 155 ------SDEATAVSLGHILNGPTLLACFGLIVTVLFMVRGIQGGIFYGMILTAIVGLISG 208
```

```
Query: 241 IVDVSNLNFGISHIGEAWTELGTTFLAAFD-GLSSLFSDSSRLPLVFMTIFAFSLSDTFD 299
           I+  +        I      L  TF  AF+  ++ +FS         + +  F   D FD
Sbjct: 209 IITYTG-----GGIVSTPPSLAPTFGQAFNIQMADVFSVQ-----FLIVVLTFLFVDFFD 258

Query: 300 TIGTFIGTGRRTGIFSQDDENALENSIGFSSKMDRALFADAIGTSIGALVGTSNTTTYVE 359
           T GT  G   + G F +D++           +  +AL AD+  TSIGA++GTS TT Y+E
Sbjct: 259 TAGTLYGVANQAG-FIKDNK---------LPRAGKALLADSSATSIGAILGTSTTTAYIE 308

Query: 360 SAAGIAEGGRTGLTAVSTAVCFLLSILLLPLVGIVPAAATAPALIIVGVMMVSSFLDVNW 419
           S+AG+A GGRTG  ++ TA  F+L++   PL+ +V    TA ALI+VG++M SS   ++W
Sbjct: 309 SSAGVAAGGRTGFASIVTAGLFVLAMFFSPLLSVVTEQVTAAALIVVGILMASSLRFIDW 368

Query: 420 SKFADALPAFFAAFFMALCYSISYGIAAAFIFYCLVKVVEGKTKDIHPIIWGATFLFIVN 479
           +K   A+P+F     M L YSI+ GIA  F+FY +  +V+G+ K++HPI++   F+F+
Sbjct: 369 TKLEIAIPSFLTVVAMPLTYSIATGIAFGFLFYPITMIVKGRGKEVHPIMYALFFVFLAY 428

Query: 480 FIILT 484
           FI L+
Sbjct: 429 FIFLS 433
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 258/488 (52%), Positives = 336/488 (67%), Gaps = 17/488 (3%)

Query: 1   MEKFFKLKEHGTTIRTEITAGLTTFFAMSYILFVNPAILSQTGMPAQGVFLATIIGAVVA 60
           MEKFFKL E+GTT+ TEI AGLTTFFAMSYILFVNP+IL   GMP+  VFLATII A ++
Sbjct: 1   MEKFFKLSENGTTVSTEIMAGLTTFFAMSYILFVNPSILGAAGMPSNAVFLATIIAAAIS 60

Query: 61  TSVMAFYANLPYAQAPGMGLNAFFTYTVVFALGYTWQEALAMVFICGLISLIITLTKVRK 120
           T +M  +AN+PYA APGMGLNAFFTYTVVFAL ++WQEALAMVFICGL ++ IT+TK RK
Sbjct: 61  TLIMGLFANVPYALAPGMGLNAFFTYTVVFALRFSWQEALAMVFICGLFNIFITVTKFRK 120

Query: 121 MIIESIPTTLKSAITAGIGTFLAYVGIKNAGFLKFSIDPGTYDVV---------GKGAAK 171
            II++IP +L+ AI  GIG F+AY+G KNA  + FSI      +V          K  A
Sbjct: 121 SIIKAIPVSLQHAIGGGIGVFVAYLGFKNANIITFSISAENIVMVNGVEPAKASAKTFAD 180

Query: 172 GLATITANSSATPGLVSFDNPAILLSLIGLSITIFFIVKGIRGGIILSILTTTLLGILMG 231
           GL  + AN    P + SF +  +LL++ GL +T   +++  RG I++ I+ TTL+GI +G
Sbjct: 181 GLLFVDANGGVVPTISSFTDSGVLLAIFGLLLTTALVIRNFRGAILIGIVATTLVGIPLG 240

Query: 232 VVKLDAINWEATNLSASFRDLKQVFGVALGEKGLISLFSNPSRLPSVLMAILAFSLTDIF 291
           +V +  +N+  +++  ++ +L   F  A    GL SLFS+ SRLP V M I AFSL+D F
Sbjct: 241 IVDVSNLNFGISHIGEAWTELGTTFLAAF--DGLSSLFSDSSRLPLVFMTIFAFSLSDTF 298

Query: 292 DTIGTLIGTGEKVGILATTGDN------HESKSLDKALYSDLIGTTFGAICGTSNVTTYV 345
           DTIGT IGTG + GI  +   +N        S  +D+AL++D IGT+ GA+ GTSN TTYV
Sbjct: 299 DTIGTFIGTGRRTGIFSQDDENALENSIGFSSKMDRALFADAIGTSIGALVGTSNTTTYV 358

Query: 346 ESAAGIGAGGRTGLTALVVAGLFAISSFFSPLVSIVPSQATAPILVIVGIMMLSNLKDIK 405
           ESAAGI  GGRTGLTA+  A  F +S    PLV IVP+ ATAP L+IVG+MM+S+  D+
Sbjct: 359 ESAAGIAEGGRTGLTAVSTAVCFLLSILLLPLVGIVPAAATAPALIIVGVMMVSSFLDVN 418

Query: 406 WDDMSEAIPAFFTSLFMGFTYSITYGIAAGFLTYTLAKVIKGQAKDIHVVLWILDILFIL 465
           W   ++A+PAFF + FM   YSI+YGIAA F+ Y L KV++G+ KDIH ++W    LFI+
Sbjct: 419 WSKFADALPAFFAAFFMALCYSISYGIAAAFIFYCLVKVVEGKTKDIHPIIWGATFLFIV 478

Query: 466 NFISLAIL 473
           NFI L IL
Sbjct: 479 NFIILTIL 486
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1111

A DNA sequence (GBSx1186) was identified in *S.agalactiae* <SEQ ID 3435> which encodes the amino acid sequence <SEQ ID 3436>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3221(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04264 GB:AP001508 unknown conserved protein [Bacillus halodurans]
 Identities = 68/147 (46%), Positives = 100/147 (67%), Gaps = 1/147 (0%)

Query: 27  MFYTQNEEELIALGQKLGTVLKSGDIVLLTGNLGAGKTTLTKGIAKGLDIKQMIKSPTYT 86
            M  TQ+ E  +A  QKL   L +GD++ L G+LGAGKT+ TKG+A GL IK+++KSPT+T
Sbjct: 5   MMITQSPEATMAFAQKLADKLLAGDVITLEGDLGAGKTSFTKGLALGLGIKRVVKSPTFT 64

Query: 87  IVREYEGRVPLYHLDVYRIGDDPDSIDLDDFLFGQGVTVIEWGELLSDNLINNYLEIVIT 146
            I+REY+GR+PLYH+DVYR+ ++ + +  D++  G GVTV+EW  L+   L    L I IT
Sbjct: 65  IIREYKGRLPLYHMDVYRLNEEEEDLGFDEYFHGDGVTVVEWASLIEGRLPPVRLAITIT 124

Query: 147 RSNQG-RQVQLEAYGHRAREIIEAIQD 172
             + +  RQ+   AYG R  E+++ + D
Sbjct: 125 HAGENERQLSFTAYGERWEEVLKELLD 151
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3437> which encodes the amino acid sequence <SEQ ID 3438>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1202(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 97/142 (68%), Positives = 122/142 (85%)

Query: 27  MFYTQNEEELIALGQKLGTVLKSGDIVLLTGNLGAGKTTLTKGIAKGLDIKQMIKSPTYT 86
            MFY++NE  L A G+ LGT L  GD+++L+G+LGAGKTTL KGIAKG+ I QMIKSPTYT
Sbjct: 1   MFYSENEYTLKAYGETLGTYLSIGDVIVLSGDLGAGKTTLAKGIAKGMGISQMIKSPTYT 60

Query: 87  IVREYEGRVPLYHLDVYRIGDDPDSIDLDDFLFGQGVTVIEWGELLSDNLINNYLEIVIT 146
            IVREYEGR+PLYHLD+YR+GDDPDSIDLDDFLFG GVTVIEWGELL + L+ +YL+I IT
Sbjct: 61  IVREYEGRLPLYHLDIYRVGDDPDSIDLDDFLFGNGVTVIEWGELLGEGLLQDYLQITIT 120

Query: 147 RSNQGRQVQLEAYGHRAREIIE 168
            + ++GRQ+ L A+G R+R+++E
Sbjct: 121 KRDKGRQLDLLAHGERSRQLLE 142
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1112

A DNA sequence (GBSx1187) was identified in *S.agalactiae* <SEQ ID 3439> which encodes the amino acid sequence <SEQ ID 3440>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1782(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35662 GB:AE001732 conserved hypothetical protein [Thermotoga maritima]
 Identities = 56/163 (34%), Positives = 94/163 (57%), Gaps = 1/163 (0%)

Query: 24  EASREEASAILEFLNTVTEETDFILHTVSNQLSLSEMETFIENTLMTKNCICLIAKLKNK 83
            EAS  +A  I+E+L  VT ETDF++        +S    +I         + ++ ++  +
Sbjct: 18  EASIWDARRIVEYLKEVTSETDFLITRPDEVYDVSTERNYIRMYRSNPGKLMIVGEINRE 77

Query: 84  VIGLITIISQSDIEIEHVGDLFIAVQKDYWGYGIGHILMEEAIEWASDNDITRRLELSVQ 143
            ++ L+T         +HVG++ I+V+K YW  GIG  ++  AIEWA  N    R++L V
Sbjct: 78  IVSLLTFTGFGRKRTKHVGEIGISVKKRYWNIGIGTRMITSAIEWARRNGFI-RIQLEVL 136

Query: 144 GRNERAIHLYQKFGFEIDGLQTRGIKRENGEFLDIYRMSKLID 186
              NERAI LY+K GFE++G++ + ++R++G F D+  M+ L+D
Sbjct: 137 KSNERAISLYRKLGFELEGIKRKAVRRDDGSFEDVLVMALLLD 179
```

There is also homology to SEQ ID 1724.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1113

A DNA sequence (GBSx1188) was identified in *S.agalactiae* <SEQ ID 3441> which encodes the amino acid sequence <SEQ ID 3442>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15582 GB:Z99122 membrane-bound protein [Bacillus subtilis]
 Identities = 108/324 (33%), Positives = 178/324 (54%), Gaps = 33/324 (10%)

Query: 5   KKITLMFSAIILTTVIALGV--YVASAYNFSTNELSKTFKDFKLAKS--KSHAIEETKPF 60
            KK TL+ + + +  ++ LG   Y    ++ + + ++   +    +K   K  +I +  PF
Sbjct: 8   KKKTLLLTILTIIGLLVLGTGGYAYYLWHKAASTVASIHESIDKSKKRDKEVSINKKDPF 67

Query: 61  SILLMGVDTGSEHRKSKWSGNSDSMILVTINPKTNKTTMTSLERDVLIKLSGPKNNGQTG 120
            S+L+MGVD     +     G +D++I +T+NPKTN T M S+ RD   K+ G       G
Sbjct: 68  SVLIMGVDERDGDK-----GRADTLIYMTVNPKTNTTDMVSIPRDTYTKIIGK------G 116

Query: 121 VEAKLNAAYASGGAEMALMTVQDLLDINVDYFMQINMQGLVDLVNAVGGITVTNKFDFPI 180
               K+N +YA GG +M + TV++ LD+ VDYF+++NM+   D+V+ +GGITV + F F
Sbjct: 117 TMDKINHSYAFGGTQMTVDTVENFLDVPVDYFVKVNMESFRDVVDTLGGITVNSTFAFSY 176

Query: 181 SIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240
               +           G   +NG++AL Y+RMR +DP GD+GRQ RQR+VIQ ++ K   +
Sbjct: 177 DGYS--------FGKGEITLNGKEALAYTRMRKEDPRGDFGRQDRQRQVIQGIINKGANI 228

Query: 241 NSISSYKKILSAVSNNMQTNIEISSKTIPNL----LAYKDSLEHIKSYQLKGEDATLSDG 296
```

```
           +SI+ +  +   V NN++TN+     T  N+      YK + +HIK ++LKG   T  +G
Sbjct: 229 SSITKFGDMFKVVENNVKTNL-----TFDNMWDIQSDYKGARKHIKQHELKG-TGTKING 282

Query: 297 GSYQILTKKHLLAVQNRIKKELDK 320
             Y    +  L  +   +K+ L+K
Sbjct: 283 IYYYQADESALSDITKELKESLEK 306
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2763> which encodes the amino acid sequence <SEQ ID 2764>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 288/436 (66%), Positives = 342/436 (78%), Gaps = 22/436 (5%)

Query: 1    MKIWKKITLMFSAIILTTVIALGVYVASAYNFSTNELSKTFKDFKLAKSKSHAIEETKPF 60
            MKI KKI LMF+AI+LTTV+ALGVY+ SAY FST ELSKTFKDF  + +KS AI++T+ F
Sbjct: 1    MKIGKKIVLMFTAIVLTTVLALGVYLTSAYTFSTGELSKTFKDFSTSSNKSDAIKQTRAF 60

Query: 61   SILLMGVDTGSEHRKSKWSGNSDSMILVTINPKTNKTTMTSLERDVLIKLSGPKNNGQTG 120
            SILLMGVDTGS  R SKW GNSDSMILVT+NPKT KTTMTSLERD L  LSGPKNN   G
Sbjct: 61   SILLMGVDTGSSERASKWEGNSDSMILVTVNPKTKKTTMTSLERDTLTTLSGPKNNEMNG 120

Query: 121 VEAKLNAAYASGGAEMALMTVQDLLDINVDYFMQINMQGLVDLVNAVGGITVTNKFDFPI 180
           VEAKLNAAYA+GGA+MA+MTVQDLL+I +D ++QINMQGL+DLVNAVGGITVTN+FDFPI
Sbjct: 121 VEAKLNAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGLIDLVNAVGGITVTNEFDFPI 180

Query: 181 SIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240
           SIA NEPEY+A V PGTHKINGEQALVY+RMRYDDPEGDYGRQKRQREVIQKVLKKILAL
Sbjct: 181 SIAENEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240

Query: 241 NSISSYKKILSAVSNNMQTNIEISSKTIPNLLAYKDSLEHIKSYQLKGEDATLSDGGSYQ 300
           +SISSY+KILSAVS+NMQTNIEISS+TIP+LL Y+D+L  IK+YQLKGEDATLSDGGSYQ
Sbjct: 241 DSISSYRKILSAVSSNMQTNIEISSRTIPSLLGYRDALRTIKTYQLKGEDATLSDGGSYQ 300

Query: 301 ILTKKHLLAVQNRIKKELDKKRSKTLKTSAILYEDYYGTTASNDSSTYSSTQENNYNTT- 359
           I+T  HLL +QNRI+ EL   +   LKT+A +YE+ YG       ST S T  NNY+++
Sbjct: 301 IVTSNHLLEIQNRIRTELGLHKVNQLKTNATVYENLYG-------STKSQTVNNNYDSSG 353

Query: 360 ---PYSEAPPSYSG-----NTTYSSETNQTTHQNYYNSSTPASNYSSNTNTGQADSSGSV 411
               YS++  SY+      +T  S+ T+Q +  + +  +TP+S+ S      ++ SSGS
Sbjct: 354 QAPSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHRPATPSSS-SDALAADESSSSGS- 411

Query: 412 NNHNGAATPNPNTGTQ 427
               G+  P  N   Q
Sbjct: 412 ----GSLVPPANINPQ 423
```

SEQ ID 3442 (GBS54) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 8; MW 48.4kDa).

The GBS54-His fusion product was purified (Figure 98A; see also Figure 194, lane 6) and used to immunise mice (lane 1+2 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 98B), FACS (Figure 98C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1114

A DNA sequence (GBSx1189) was identified in *S.agalactiae* <SEQ ID 3443> which encodes the amino acid sequence <SEQ ID 3444>. This protein is predicted to be Vesl-1L. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.44    Transmembrane    3 -  19 (   3 -  19)

----- Final Results -----
              bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3445> which encodes the amino acid sequence <SEQ ID 3446>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 42/98 (42%), Positives = 64/98 (64%)

Query: 1  MKIGRLIALGLVSLGALELYKNRKTIKDSYQNTKNETDSAKLKLERIKNDLAIISQEKEK 60
           MK+  +IA+GL+S  A + Y+ R TIK+     ++  D+A+L L+ IK +L +I  + +
Sbjct: 1  MKVKTVIAVGLLSFTAYKAYQKRCTIKELLSISRQAKDAAQLDLDNIKANLDLIHSQGKV 60

Query: 61 IRLISQELNHKFQVFNKDIQPRLEEINQRMAKYQEKDE 98
           I+ ISQ+L HK++ FN++ Q  L EI  RMAKYQE  E
Sbjct: 61 IQNISQDLAHKWRYFNQETQAHLTEIQNRMAKYQEDSE 98
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1115

A DNA sequence (GBSx1190) was identified in *S.agalactiae* <SEQ ID 3447> which encodes the amino acid sequence <SEQ ID 3448>. This protein is predicted to be Hit-like protein involved in cell-cycle regulation (hit). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2694(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04908 GB:AP001511 Hit-like protein involved in cell-cycle
           regulation [Bacillus halodurans]
 Identities = 74/137 (54%), Positives = 95/137 (69%), Gaps = 2/137 (1%)

Query: 3   NCIFCKIISGEIPSSKVYEDDEVLAFLDITQTTTGHTLLIPKKHVRNVLEMDEKTAQITF 62
           NCIFCKII+GEIPS+ VYEDD V AFLDI+Q T GHTL+IPK H RNV E+ E+ A   F
Sbjct: 6   NCIFCKIIAGEIPSATVYEDDHVYAFLDISQVTKGHTLVIPKVHKRNVFELSEEIASSLF 65

Query: 63  ERLPKVARAVQAATKAKGMNIINNNEEIAGQTVFHAHVHLVPRFDESDGIKIHYTTHEPD 122
             +PK++RA+  A +  GMNI+NNN E AGQTVFH H+HL+PR+ E DG    +  H
Sbjct: 66  AAVPKISRAINDAFQPIGMNIVNNNGEAAGQTVFHYHLHLLPRYGEGDGYGAVWKDHSSQ 125

Query: 123 F--EALAKLAKEIRKEI 137
           +  + L  L+  IR+ +
Sbjct: 126 YSGDDLQVLSSSIREHL 142
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3449> which encodes the amino acid sequence <SEQ ID 3450>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0125(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 97/137 (70%), Positives = 117/137 (84%)

Query: 1   MDNCIFCKIISGEIPSSKVYEDDEVLAFLDITQTTTGHTLLIPKKHVRNVLEMDEKTAQI 60
           M+NCIFC II G+IPSSKVYED++VLAFLDI+QTT GHTL+IPK+HVRN+LEM  +TA
Sbjct: 1   MENCIFCSIIQGDIPSSKVYEDEQVLAFLDISQTTKGHTLVIPKQHVRNLLEMTAETASH 60

Query: 61  TFERLPKVARAVQAATKAKGMNIINNNEEIAGQTVFHAHVHLVPRFDESDGIKIHYTTHE 120
            F R+PK+ARA+Q+AT A  MNIINNNE +AGQTVFHAHVHLVPR++E DGI I YTTHE
Sbjct: 61  LFARIPKIARAIQSATGATAMNIINNNEALAGQTVFHAHVHLVPRYNEEDGISIQYTTHE 120

Query: 121 PDFEALAKLAKEIRKEI 137
           PDF  L KLA++I +E+
Sbjct: 121 PDFPVLEKLARQINQEV 137
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1116

A DNA sequence (GBSx1191) was identified in *S.agalactiae* <SEQ ID 3451> which encodes the amino acid sequence <SEQ ID 3452>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10923> which encodes amino acid sequence <SEQ ID 10924> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3452 (GBS87) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 3; MW 19.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 10; MW 44kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1117

A DNA sequence (GBSx1192) was identified in *S.agalactiae* <SEQ ID 3453> which encodes the amino acid sequence <SEQ ID 3454>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.53    Transmembrane  143 - 159 ( 141 - 161)

----- Final Results -----
               bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9563> which encodes amino acid sequence <SEQ ID 9564> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12844 GB:Z99109 ABC transporter (ATP-binding protein)
           [Bacillus subtilis]
 Identities = 137/242 (56%), Positives = 181/242 (74%)

Query: 1   MTMLKIENVTGGYVNIPVLKNISFEVNDGELVGLIGLNGAGKSTTINEIIGILRPYQGDI 60
            M++L ++++TGGY   PVLKN+SF +   ++VGLIGLNGAGKSTTI  IIG++ P++G I
Sbjct: 1   MSLLSVKDLTGGYTRNPVLKNVSFTLEPNQIVGLIGLNGAGKSTTIRHIIGLMDPHKGSI 60

Query: 61  TIDGISLEADQELYRKKIGFIPETPSLYEELTLREHLETVAMAYDIATDEVMARAQKLLE 120
             ++G +   D E YR +  +IPETP LYEELTL EHLE  AMAY ++ + +  R   LL+
Sbjct: 61  ELNGKTFAEDPEGYRSQFTYIPETPVLYEELTLMEHLELTAMAYGLSKETMEKRLPPLLK 120

Query: 121 MFRLTDKLDWFPMHFSKGMKQKVMIICAFVVSPSLFIVDEPFLGLDPLAISDLINLLAEE 180
             FR+  +L WFP HFSKGMKQKVMI+CAF+  P+L+I+DEPFLGLDPLAI+ L+  + E
Sbjct: 121 EFRMEKRLKWFPAHFSKGMKQKVMIMCAFLAEPALYIIDEPFLGLDPLAINALLERMNEA 180

Query: 181 KAKGKSILMSTHVLDSAEKMCDRFVILHKGEIRAVGTLEELRAIFGDSNANLNDIYIALT 240
            K  G S+LMSTH+L +AE+ CD F+ILH GE+RA GTL ELR  FG  +A L+D+Y+ LT
Sbjct: 181 KKGGASVLMSTHILATAERYCDSFIILHNGEVRARGTLSELREQFGMKDAALDDLYLELT 240

Query: 241 KE 242
            KE
Sbjct: 241 KE 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3455> which encodes the amino acid sequence <SEQ ID 3456>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.04    Transmembrane  141 - 157 ( 139 - 158)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.3017(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12844 GB:Z99109 ABC transporter (ATP-binding protein)
           [Bacillus subtilis]
 Identities = 139/241 (57%), Positives = 189/241 (77%)

Query: 1   MLNIKNLTGGYHNIPVLNDVSFSVDNGELVGLIGLNGAGKSTTINEIIGFLKPYQGSISI 60
           +L++K+LTGGY   PVL +VSF+++  ++VGLIGLNGAGKSTTI  IIG + P++GSI +
Sbjct: 3   LLSVKDLTGGYTRNPVLKNVSFTLEPNQIVGLIGLNGAGKSTTIRHIIGLMDPHKGSIEL 62

Query: 61  DGLTLAENAVAYRQKIGFIPETPSLYEELTLSEHINTVAMAYDIDLEVAQKRAQPFLEMF 120
           +G T AE+   YR +  +IPETP LYEELTL EH+   AMAY +  E  +KR  P L+ F
Sbjct: 63  NGKTFAEDPEGYRSQFTYIPETPVLYEELTLMEHLELTAMAYGLSKETMEKRLPPLLKEF 122

Query: 121 RLTDKLEWFPVNFSKGMKQKVMIICAFVIDPSLFILDEPFLGLDPLAISDLIQTLEVEKA 180
           R+  +L+WFP +FSKGMKQKVMI+CAF+ +P+L+I+DEPFLGLDPLAI+ L++ +   K
Sbjct: 123 RMEKRLKWFPAHFSKGMKQKVMIMCAFLAEPALYIIDEPFLGLDPLAINALLERMNEAKK 182

Query: 181 KGKSILMSTHVLDSAERMCDRFVILHHGQVRAQGTLADLQEAFGDRSASLNDIYLALTKED 241
            G S+LMSTH+L +AER CD F+ILH+G+VRA+GTL++L+E FG + A+L+D+YL LTKED
Sbjct: 183 GGASVLMSTHILATAERYCDSFIILHNGEVRARGTLSELREQFGMKDAALDDLYLELTKED 243
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 181/240 (75%), Positives = 208/240 (86%)

Query: 3   MLKIENVTGGYVNIPVLKNISFEVNDGELVGLIGLNGAGKSTTINEIIGILRPYQGDITI 62
           ML I+N+TGGY NIPVL ++SF V++GELVGLIGLNGAGKSTTINEIIG L+PYQG I+I
Sbjct: 1   MLNIKNLTGGYHNIPVLNDVSFSVDNGELVGLIGLNGAGKSTTINEIIGFLKPYQGSISI 60

Query: 63  DGISLEADQELYRKKIGFIPETPSLYEELTLREHLETVAMAYDIATDEVMARAQKLLEMF 122
           DG++L  +   YR+KIGFIPETPSLYEELTL EH+ TVAMAYDI  +    RAQ  LEMF
Sbjct: 61  DGLTLAENAVAYRQKIGFIPETPSLYEELTLSEHINTVAMAYDIDLEVAQKRAQPFLEMF 120

Query: 123 RLTDKLDWFPMHFSKGMKQKVMIICAFVVSPSLFIVDEPFLGLDPLAISDLINLLAEEKA 182
           RLTDKL+WFP++FSKGMKQKVMIICAFV+ PSLFI+DEPFLGLDPLAISDLI  L  EKA
Sbjct: 121 RLTDKLEWFPVNFSKGMKQKVMIICAFVIDPSLFILDEPFLGLDPLAISDLIQTLEVEKA 180

Query: 183 KGKSILMSTHVLDSAEKMCDRFVILHKGEIRAVGTLEELRAIFGDSNANLNDIYIALTKE 242
           KGKSILMSTHVLDSAE+MCDRFVILH G++RA GTL +L+  FGD +A+LNDIY+ALTKE
Sbjct: 181 KGKSILMSTHVLDSAERMCDRFVILHHGQVRAQGTLADLQEAFGDRSASLNDIYLALTKE 240
```

SEQ ID 3454 (GBS353) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 2; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 6; MW 55kDa).

GBS353-GST was purified as shown in Figure 216, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1118

A DNA sequence (GBSx1193) was identified in *S.agalactiae* <SEQ ID 3457> which encodes the amino acid sequence <SEQ ID 3458>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1475(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1119

A DNA sequence (GBSx1194) was identified in *S.agalactiae* <SEQ ID 3459> which encodes the amino acid sequence <SEQ ID 3460>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.68   Transmembrane   57 -  73 (  50 -  80)
   INTEGRAL    Likelihood = -8.49   Transmembrane  122 - 138 ( 103 - 152)
   INTEGRAL    Likelihood = -6.58   Transmembrane  319 - 335 ( 308 - 337)
   INTEGRAL    Likelihood = -4.99   Transmembrane  252 - 268 ( 249 - 273)
   INTEGRAL    Likelihood = -4.19   Transmembrane  104 - 120 ( 103 - 121)
   INTEGRAL    Likelihood = -3.50   Transmembrane  231 - 247 ( 229 - 248)
   INTEGRAL    Likelihood = -1.91   Transmembrane  298 - 314 ( 298 - 314)
   INTEGRAL    Likelihood = -1.44   Transmembrane   28 -  44 (  27 -  44)

----- Final Results -----
               bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12845 GB:Z99109 ABC transporter (membrane protein) [Bacillus subtilis]
 Identities = 101/409 (24%), Positives = 187/409 (45%), Gaps = 76/409 (18%)

Query: 1   MKKLFNKRRSLFLTQNSKYLRYVFNDHFVLVLMFLSGFLLYQYSQLLKDFPKTHWPIIVI 60
           M  ++  R    + +   Y++Y+ NDH V+VL+F        YS+ ++D P  H+P   +
Sbjct: 4   MLDIWQSRLQEHIKETRTYMKYMLNDHLVIVLIFFLAGAASWYSKWIRDIP-AHFPSFWV 62

Query: 61  VSIIILMLLAMGGIASYLEPADKQFLLIKEEAIKEIINSAKKRTYI-------------- 106
           ++++  ++L    + + L+ AD  FLL  E  ++  +  A   +Y+
Sbjct: 63  MAVLFSLVLTSSYVRTLLKEADLVFLLPLEAKMEPYLKQAFVYSYVSQLFPLIALSIVAM 122

Query: 107 --FWLVIQTLFLVLISPILIKLGL------------------------------------ 128
             ++ V     LV  + + ++L L
Sbjct: 123 PLYFAVTPGASLVSYAAVFVQLLLLKAWNQVMEWRTTFQNDRSMKRMDVIIRFAANTLVL 182

Query: 129 -----SVFMITLLIFGLGIIKWLVITYKVKVFYNNQNLNWDAAINHEQERKQSILKFFSL 183
                SV+M  LL++ +  + +L ++   K     +   W++ I  E  RKQ   +  +L
Sbjct: 183 YFVFQSVYMYALLVYVIMAVLYLYMSSAAK----RKTFKWESHIESELRRKQRFYRIANL 238

Query: 184 FTNVKGISTSVKRRSFLDGILKLISKTPSRLWTNLFVRAFLRSSDYLGLTIRLVTLNILS 243
           FT+V  +    KRR++LD +L+L+     + +  +F RAFLRSSDYLG+ +RL  +  L
Sbjct: 239 FTDVPHLRKQAKRRAYLDFLLRLVPFEQRKTFAYMFTRAFLRSSDYLGILVRLTIVFALI 298

Query: 244 VIFVNETYLALALAFVFN-YLLLFQLLALGHHFDYQYMNQLYPVRLNAKASQLKGFLRVL 302
           +++V+ + L  A+  VF  ++   QLL L  HFD+  + +LYPV+   K ++LK +  +L
Sbjct: 299 IMYVSASPLIAAVLTVFAIFITGIQLLPLFGHFDHLALQELYPVQ---KETKLKSYFSLL 355

Query: 303 SYAVTVIDSI----------LIRELKPVILLIVLMLIVTEYYIPYKIKK 341
             A+++   +          L   L  +I   VL+ +V   Y+  ++KK
```

```
Sbjct: 356 KTALSIQALLMSVASAYAAGLTGFLYALIGSAVLIFVVLPAYMTTRLKK 404
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3461> which encodes the amino acid sequence <SEQ ID 3462>. Analysis of this protein sequence reveals the following:

```
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.91   Transmembrane  126 - 142 ( 119 - 151)
   INTEGRAL    Likelihood = -9.77   Transmembrane  320 - 336 ( 311 - 339)
   INTEGRAL    Likelihood = -6.37   Transmembrane   59 -  75 (  53 -  79)
   INTEGRAL    Likelihood = -4.94   Transmembrane   28 -  44 (  22 -  47)
   INTEGRAL    Likelihood = -4.73   Transmembrane  250 - 266 ( 249 - 273)
   INTEGRAL    Likelihood = -4.04   Transmembrane  231 - 247 ( 229 - 248)
   INTEGRAL    Likelihood = -3.19   Transmembrane  298 - 314 ( 295 - 315)
   INTEGRAL    Likelihood = -2.28   Transmembrane  103 - 119 ( 103 - 119)

----- Final Results -----
               bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12845 GB:Z99109 ABC transporter (membrane protein) [Bacillus subtilis]
 Identities = 96/403 (23%), Positives = 173/403 (42%), Gaps = 78/403 (19%)

Query: 1   MKALFLKRRQDFQKQQNKYLRYVLNDHFVLVLMFLLGFAMVQYGQLLN----HFPT---- 52
           M  ++  R Q+  K+   Y++Y+LNDH V+VL+F L  A    Y + +    HFP+
Sbjct: 4   MLDIWQSRLQEHIKETRTYMKYMLNDHLVIVLIFFLAGAASWYSKWIRDIPAHFPSFWVM 63

Query: 53  -------------NHLPIQVCLGILIPLLLSM---------------------------- 71
                          L  +  L  L+PL   M
Sbjct: 64  AVLFSLVLTSSYVRTLLKEADLVFLLPLEAKMEPYLKQAFVYSYVSQLFPLIALSIVAMP 123

Query: 72  --------GSIATYLEEADQHFLLPKEEEVISYI------KQAERLSFLLWGTLQTAVLL 117
                    S+ +Y     Q  LL    +V+ +       +  +R+  ++     T VL
Sbjct: 124 LYFAVTPGASLVSYAAVFVQLLLLKAWNQVMEWRTTFQNDRSMKRMDVIIRFAANTLVLY 183

Query: 118 FLYPIFRRLGLSLFIFIILVLILLALKRVVLSRKTRYFLRGNRLDWAKAVAFESNRKQSI 177
           F++        S++++ +LV +++A+  + +S   +         W   +  E  RKQ
Sbjct: 184 FVFQ-------SVYMYALLVYVIMAVLYLYMSSAAKR----KTFKWESHIESELRRKQRF 232

Query: 178 LKFYSLFTTVKGISTKVKERTYLNPLLKLVKQTPSNLWLSLYARAFLRSSDYLGLFLRLM 237
            +  +LFT V  +  + K R YL+ LL+LV     +    ++ RAFLRSSDYLG+ +RL
Sbjct: 233 YRIANLFTDVPHLRKQAKRRAYLDFLLRLVPFEQRKTFAYMFTRAFLRSSDYLGILVRLT 292

Query: 238 LLSSLSVFFIHNLYLSVSLALIFN-YLVVFQLLSLYYHYDYHYMTSLYPENSRSKKKNML 296
           ++ +L + ++    L  ++  +F  ++   QLL L+ H+D+  +  LYP    +K K+
Sbjct: 293 IVFALIIMYVSASPLIAAVLTVFAIFITGIQLLPLFGHFDHLALQELYPVQKETKLKSYF 352

Query: 297 SFLR-GLSFLMLIVNMLCCSSAPKA--LILIVGMVFIACIYLP 336
           S L+  LS   L++++    +A     L  ++G   +  + LP
Sbjct: 353 SLLKTALSIQALLMSVASAYAAGLTGFLYALIGSAVLIFVVLP 395
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 170/344 (49%), Positives = 237/344 (68%)

Query: 1   MKKLFNKRRSLFLTQNSKYLRYVFNDHFVLVLMFLSGFLLYQYSQLLKDFPKTHWPIIVI 60
           MK LF KRR  F  Q +KYLRYV NDHFVLVLMFL GF + QY QLL  FP  H PI V
Sbjct: 1   MKALFLKRRQDFQKQQNKYLRYVLNDHFVLVLMFLLGFAMVQYGQLLNHFPTNHLPIQVC 60

Query: 61  VSIIILMLLAMGGIASYLEPADKQFLLIKEEAIKEIINSAKKRTYIFWLVIQTLFLVLIS 120
           + I+I +LL+MG IA+YLE AD+ FLL KEE +   I  A++ +++ W  +QT  L+ +
Sbjct: 61  LGILIPLLLSMGSIATYLEEADQHFLLPKEEEVISYIKQAERLSFLLWGTLQTAVLLFLY 120

Query: 121 PILIKLGLSVFMITLLIFGLGIIKWLVITYKVKVFYNNQNLNWDAAINHEQERKQSILKF 180
```